

THÈSE

présentée

A LA FACULTÉ DES SCIENCES ET DES TECHNIQUES DE
L'UNIVERSITÉ DE FRANCHE-COMTÉ

BESANCON

pour obtenir le

DIPLÔME DE DOCTEUR EN MATHÉMATIQUES

par

Abderrahim MESSAOUDI

**SUR LA DÉCOMPOSITION INCOMPLÈTE
DE CERTAINES CLASSES DE MATRICES :
ALGORITHMES ITÉRATIFS ASSOCIÉS**

Soutenue le 4 Juillet 1983, devant la Commission d'Examen :

M. ROBERT François

Président

MM. BENILAN Philippe

JOLY Pascal

LESAINTE Pierre

MIELLOU Jean-Claude

Examineurs

THÈSE

présentée

A LA FACULTÉ DES SCIENCES ET DES TECHNIQUES DE
L'UNIVERSITÉ DE FRANCHE-COMTÉ

pour obtenir le

DIPLÔME DE DOCTEUR EN MATHÉMATIQUES

par

Abderrahim MESSAOUDI

**SUR LA DÉCOMPOSITION INCOMPLÈTE
DE CERTAINES CLASSES DE MATRICES :
ALGORITHMES ITÉRATIFS ASSOCIÉS**

Soutenue le 4 Juillet 1983, devant la Commission d'Examen :

M. ROBERT François

Président

MM. BENILAN Philippe

JOLY Pascal

LESAINT Pierre

MIELLOU Jean-Claude

Examineurs

UNIVERSITÉ DE FRANCHE-COMTÉ BESANÇON**Président :** Monsieur le Professeur Jacques ROBERT**Présidents honoraires :** MM. THIEBAUT, LÉVÊQUE**FACULTÉ DES SCIENCES ET DES TECHNIQUES****DOYEN**M. le Professeur Jean BULABOIS
Physique Générale-optique**ASSEESSEURS**M. le Professeur Bernard MILLET
Botanique
M. HEROLD Jean-Pierre, Maître-Assistant
Physiologie animale**DOYENS HONORAIRES**MM. GLANGEAUD, JACQUEMAIN, CHATELET, THIÉBAUT,
BERNARD, GALATRY**PROFESSEURS HONORAIRES**MM. BENNETON, CHATELET, MESNAGE, PERROT,
PLUVINAGE, QUANTIN, REAL, THIRY, TRILLAT,
TRONCHET, UEBERSFELD**PROFESSEURS**

MM. ALBIGNAC Roland	: Biologie anim. et Ecologie	LAUDE Bernard	: Chimie organique
BANTEGNIE Robert	: Mathématique	LESAIN Pierre	: Mathématiques I
BENILAN Philippe	: Mathématiques	LHOTE François	: Automatique
BERNARD Jean	: Chimie Physique	MANIA Jacky	: Géologie Structurale et appliquée
BIDAULT Michel	: Taxonomie expérimentale		
Mme BRIDE Michèle	: Zoologie Embryologie	MERIGOUX Henri	: Cristallographie et Synthèses minérales
BROQUET Paul	: Géologie		
BRUCKERT Sylvain	: Ecologie végétale	MIELLOU Jean-Claude	: Analyse numérique
BULABOIS Jean	: Physique Gle Optique	MILLET Bernard	: Botanique
CAPODANNO Pierre	: Mécanique théorique	MONTAGNER Hubert	: Psychophysiologie
CERUTTI Ernest	: Chimie appliquée	MOREELS Guy	: Astronomie
CHALEAT Raymond	: Mécanique appliquée	OLIVIER Marcel	: Electronique quantique
CHAMBAUDET Alain	: Chimie	OYTANA Claude	: Mécanique appliquée
CHARDON Jean-Claude	: Spectroscopie	PAGETTI Claude	: Corrosion et traitements de surface
CHAUVE Pierre	: Géologie minéralogie		
COUGNARD Jean	: Mathématiques	PARIZET Jean	: Mathématiques
DEVIN Claude	: Chmie 1er cycle	POTIER Robert	: Physique des solides
DOBREMEZ Jean-François	: Biologie	RANGHEARD Yves	: Géologie
DUBOUCHET Jacques	: Physiologie végétale	RIPPLINGER Jean	: Physiologie animale
GALATRY Louis	: Physique moléculaire	ROBERT Daniel	: Physique
GAUDEMER Yves	: Biochimie	ROBERT Guy	: Electrochimie
GIRARDET Claude	: Physique moléculaire	ROBERT Jacques	: Mathématiques I
GOMOT Lucien	: Zoologie-Embryologie	STRICKER Christophe	: Mathématiques
GOUARNE René	: Mathématiques	THEOBALD Jean-Gérard	: Spectroscopie hertzienne
GRAS Georges	: Mathématiques	THIEBAUT Jean	: Petrogr. minéralogie
GREMILLARD Jean	: Mathématiques	TREHEL Michel	: Informatique
JOLICARD Georges	: Physique moléculaire	VERNEAUX Jean	: Zoologie-Embryologie
KARCHE Jean-Paul	: Pétrographie Minéralogie	VIENTOT Jean-Charles	: Physique Gle Optique
LALLEMENT Gérard	: Mécanique appliquée	VIGOUREUX Jean-Marie	: Physique
		WEIL Michel	: Mathématiques

Responsable des services administratifs: M. Jean-Marie RUNGE**Professeurs des Instituts Universitaires de Technologie dont les activités de Recherche sont rattachées à la Faculté des Sciences et Techniques de Besançon****I.U.T. de Besançon**MM. GENTYL Claude : Spécialité Chimie
ROUYER Bernard : Spécialité Mécanique**I.U.T. de Belfort**KAUFFMANN Jean-Marie : Spécialité Electrotechnique
MANDRET Gérard : Spécialité Electronique
PRENEL Jean-Pierre : Spécialité Physique
STAMON Georges : Spécialité Informatique
RONDOT Daniel : Spécialité Métallurgie**ECOLE NATIONALE SUPÉRIEURE DE MICROMÉCANIQUE ET MICROTECHNIQUE****DIRECTEUR**M. le Professeur Raymond CHALEAT
Mécanique Appliquée**DIRECTEUR HONORAIRE**

M. le Professeur MESNAGE

MM. ANDRE Pierre : Informatique Industrielle
AVRAMESCO André : Mathématiques
BESSON Raymond : Electronique
FAURE Alain : AutomatiqueHAUDEN Daniel : Microélectronique
Mme LOSCO Lucette : Mécanique
ROQUES-CARMES : Physico-Chimie des Matériaux**Responsable des services administratifs :** Madame Jocelyne HAUTON



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41546621_0024

Je tiens particulièrement à remercier :

Monsieur F. ROBERT, Professeur à l'Université de Grenoble, pour l'intérêt qu'il porte à ce travail et pour l'honneur qu'il me fait de bien vouloir présider ce jury de thèse. Monsieur F. ROBERT a contribué à ma formation, il m'a initié à la recherche et m'a proposé le thème de ce travail. Je lui suis très reconnaissant,

Monsieur J.-C. MIELLOU, Professeur à l'Université de Besançon, qui a bien voulu accepter de diriger ce travail. Je veux aussi lui exprimer ma sincère gratitude, pour la confiance qu'il m'a toujours témoignée, pour sa disponibilité à mon égard et pour ses nombreux conseils et encouragements,

Monsieur P. JOLY, Ingénieur au C.N.R.S., qui a manifesté son intérêt pour ce travail et a accepté de le juger,

Monsieur P. LESAIN, Professeur à l'Université de Besançon, qui a bien voulu accepter de faire partie du jury,

Monsieur Ph. BENILAN, Professeur à l'Université de Besançon, qui a bien voulu accepter de juger ce travail et de faire partie du jury,

Mademoiselle C. PAGANI qui, avec compétence, soin et bonne humeur, a assuré la dactylographie de cette thèse, ainsi que tout le personnel du service de reproduction.

★ ★
★

INTRODUCTION

ERRATA

=====

* Les théorèmes 2.1 et 2.2 sont les mêmes pages 20 et 21.

* Remplacer "matrice anti-symétrique et définie positive" par
matrice dont les éléments diagonaux sont positifs et les parties
inférieures et supérieures sont anti-symétriques - pages 124
et 133.

★ ★
★

$$\frac{1 + \rho((LU)^{-1}R)}{1 - \rho((LU)^{-1}R)} \leq \frac{2}{1 - \rho((LU)^{-1}R)}$$

INTRODUCTION

L'objet de ce travail est l'étude des résolutions numériques de systèmes linéaires : (*) $Ax = b$, où A est une matrice carrée non singulière, par des méthodes itératives indirectes et des algorithmes itératifs.

Pour les méthodes indirectes, on utilise l'élimination de Gauss pour décomposer la matrice A sous la forme :

$$A = LU - R$$

et on utilise la méthode de point fixe associée à cette décomposition :

$$LUx = Rx + b.$$

Pour une M -matrice, Meijerink et Van der Vorst [15] ont proposé un procédé pour une telle décomposition et ont montré que le rayon spectral $\rho((LU)^{-1}R) < 1$.

Au premier chapitre, nous rappelons quelques propriétés des M -matrices et le procédé utilisé par Meijerink et Van der Vorst pour la décomposition incomplète de telles matrices.

Au cours de l'élaboration de ce travail, d'autres études ([13], [14], [24], [29], [12]) ont été faites pour l'application de ce procédé à d'autres classes de matrices (soit définies positives, soit H -matrices).

Au second chapitre, après un rappel de quelques propriétés des H -matrices, nous explicitons le cas de la décomposition incomplète des H -matrices d'une façon plus précise que ce qui existe dans la littérature. De plus, on démontre que $\rho((LU)^{-1}R) < 1$ pour de telles matrices.

Ce résultat donne un renseignement sur le conditionnement de la matrice : $I - (LU)^{-1}R$ que l'on peut alors estimer par :

$$\frac{1 + \rho((LU)^{-1}R)}{1 - \rho((LU)^{-1}R)} < \frac{2}{1 - \rho((LU)^{-1}R)}.$$

Ensuite, on énonce quelques nouvelles caractérisations des H-matrices qui ont un rapport avec ce résultat et leur décomposition incomplète.

Au troisième chapitre, on reprend une idée de Manteuffel [14] pour proposer une méthode qui généralise la décomposition incomplète à des matrices quelconques ; cette méthode consiste à transformer la matrice initiale en une H-matrice et à appliquer la décomposition sur cette dernière.

En pratique, les décompositions incomplètes se combinent avec des méthodes d'accélération de la convergence de type gradient conjugué [1], [2], [5], [6], [8], [10], [12], [16], [20], [26] double-gradient conjugué [32]. Un algorithme dit d'orthogonalisation - minimisation a été proposé par Vinsome [30] pour la résolution du système (*) présentant un cas particulier : la matrice A est une M-matrice symétrique ; l'étude de la convergence et du comportement de cet algorithme n'a pas été faite.

Au quatrième chapitre, nous reprenons l'étude de l'algorithme d'orthogonalisation - minimisation (ORTHOMIN) ; nous allons proposer une forme généralisée de cet algorithme qui s'appliquera à tout problème même non symétrique. Ensuite, nous donnons un résultat sur les conditions de sa convergence. Nous rappelons aussi l'algorithme du gradient conjugué généralisé et l'algorithme du double-gradient conjugué.

Au cinquième chapitre qui est consacré à des essais numériques sur trois types de problèmes différents, on procède à une comparaison de ces algorithmes appliqués avec des décompositions incomplètes.

- CHAPITRE 1 -
=====

DECOMPOSITIONS INCOMPLETES
DES M-MATRICES

Introduction :

Dans ce chapitre, nous donnerons quelques propriétés des M-matrices qui nous serviront dans le chapitre suivant ; nous rappellerons aussi le procédé utilisé par Meijerink - Van Der Vorst [15] pour la démonstration du théorème des décompositions incomplètes des M-matrices ; ce procédé, nous l'utiliserons pour la démonstration de l'extension de ce théorème à des H-matrices.

- CHAPITRE 1 -

1 - DEFINITION ET PROPRIETES DES M-MATRICES

Soit $A = (a_{ij})$ une matrice carrée d'ordre n à coefficients réels.
On notera $I = \{1, 2, \dots, n\}$ et on posera :
 $A = D - E - F$ avec $D = (a_{ii})$ pour $i \in I$, $-E = (a_{ij})_{i > j}$ pour $i, j \in I$
et $-F = (a_{ij})_{j > i}$ pour $i, j \in I$.
On notera aussi $|A| = (|a_{ij}|)$ et θ la matrice nulle.

Définition 1.1 :

On dira que la matrice A est non négative (resp. positive) et on notera $A \geq \theta$ (resp. $A > \theta$) si $a_{ij} \geq 0 \quad \forall i, j \in I$ (resp. $a_{ij} > 0$).

Définition 1.2 :

On dira que A est une Z -matrice si :

$$\begin{cases} a_{ii} \geq 0 & \forall i \in I \\ a_{ij} \leq 0 & \text{pour } i \neq j \text{ et } i, j \in I. \end{cases}$$

Définition 1.3 :

On dira que A est une M -matrice si :

A est une Z -matrice

A est non singulière et $A^{-1} \geq \theta$.

Une M -matrice ainsi définie est donc non singulière.

Définition 1.4 :

On désignera par $N(A)$ la Z-matrice associée à A (ou la matrice de comparaison) définie par : $N(A) = |D| - |E| - |F|$.
Si $a_{ii} \neq 0$ pour tout $i \in I$, on désignera par J_A la matrice de Jacobi associée à A définie par : $J_A = I - D^{-1}A$.

Définition 1.5 :

On définit le rayon spectral de A , $\rho(A)$, par :
 $\rho(A) = \max |\lambda_i|$, où les λ_i sont les valeurs propres de A .

Proposition 1.1 :

Soit $A = (a_{ij})$ une Z-matrice. Des conditions nécessaires et suffisantes pour que A soit une M-matrice sont :

- 1 - Les éléments diagonaux de A sont positifs ($D > 0$)
- 2 - $J_A = I - D^{-1}A \geq \theta$ et $\rho(J_A) < 1$.

Preuve : cf. Varga [27]

On pose $I_k = \{1, 2, \dots, k\}$ avec $k = 1, 2, \dots, n$.

On définit les matrices mineures principales $(A_k)_{k=1, \dots, n}$ de A par :

$A_k = (a_{ij})$ pour $i, j \in I_k$.

On a évidemment $A_n = A$.

Théorème 1.1 :

Soit A une Z-matrice. Les assertions suivantes sont équivalentes :

- 1 - A est une M-matrice
- 2 - Tous les mineurs principaux de A sont positifs
- 3 - Il existe L et U matrices triangulaires respectivement inférieure et supérieure à diagonales positives tel que $A = LU$.

Preuve : cf. Bermann - Plemmons [3]

De ce théorème découle le corollaire suivant.

Corollaire 1.1 :

Si A est une M -matrice, alors toutes les matrices mineures principales de A sont aussi des M -matrices.

Théorème 1.2 :

Soient $A = (a_{ij})$ une M -matrice et $B = (b_{ij})$ la matrice telles que : $a_{ij} \leq b_{ij} \leq 0$ pour $i \neq j$ et $0 < a_{ii} \leq b_{ii}$; alors B est aussi une M -matrice et on a $A^{-1} \geq B^{-1} \geq 0$.

Preuve :

Pour montrer que B est une M -matrice, on va utiliser la proposition 1.1 :

$$D_B = (b_{ii}) \geq (a_{ii}) = D_A > 0$$

$$J_B = I - D_B^{-1} B \geq 0$$

$$\text{Or } -\frac{b_{ij}}{b_{ii}} \leq -\frac{a_{ij}}{a_{ii}} \text{ ce qui donne matriciellement}$$

$$\theta \leq J_B \leq J_A \text{ d'où } \rho(J_B) \leq \rho(J_A) < 1$$

et B est donc une M -matrice.

Comme $A = D_A (I - J_A)$ et $B = D_B (I - J_B)$ sont des M -matrices alors :

$$A^{-1} = (I - J_A)^{-1} D_A^{-1} \geq 0 \quad \text{avec } (I - J_A)^{-1} \geq 0 \quad \text{et } D_A^{-1} > 0$$

$$B^{-1} = (I - J_B)^{-1} D_B^{-1} \geq 0 \quad \text{avec } (I - J_B)^{-1} \geq 0 \quad \text{et } D_B^{-1} > 0.$$

$$\text{Or comme } \theta \leq J_B \leq J_A \quad \text{et} \quad \rho(J_B) \leq \rho(J_A) < 1$$

$$\text{alors } (I - J_B)^{-1} \leq (I - J_A)^{-1}.$$

$$\text{On a aussi } D_A^{-1} \geq D_B^{-1} \text{ d'où}$$

$$A^{-1} = (I - J_A)^{-1} D_A^{-1} \geq (I - J_B)^{-1} D_B^{-1} = B^{-1}.$$

on a :

$$B = L A.$$

D'après les formules (1.1), on a :

$$b_{ij} = a_{ij} - \frac{a_{ik} a_{kj}}{a_{kk}} \leq a_{ij} \leq 0 \quad \text{pour } i \neq j \quad \text{et } j \neq k$$

$$\text{et} \quad b_{ik} = 0 \quad \text{pour } i \neq k$$

$$\text{donc} \quad b_{ij} \leq 0 \quad \text{pour } i \neq j.$$

On a aussi pour $i = k+1, \dots, n$:

$$b_{ii} = a_{ii} - \frac{a_{ik} a_{ki}}{a_{kk}} > 0$$

car sinon on contredirait le fait que toutes les matrices mineures principales de A sont des M -matrices et ont leurs déterminants positifs

$$\text{donc} \quad b_{ii} > 0 \quad \text{pour } i \in I$$

et par conséquent B est une Z -matrice.

Considérons maintenant les matrices mineures principales de B :

$$B_m = (L A)_m = L_m A_m \quad \text{pour } m \in I$$

$$\text{or} \quad \det L_m = 1 \quad \text{d'où :}$$

$$\det B_m = \det L_m \cdot \det A_m$$

$$= \det A_m > 0$$

donc B , une Z -matrice dont tous ses mineurs principaux sont positifs, est une M -matrice.

$$\text{Montrons que : } A^{-1} \geq B^{-1}.$$

Remarquons d'abord que :

$$L^{-1} = 2I - L \quad \text{et} \quad L - I \geq \theta.$$

On a alors :

$$\begin{aligned} A^{-1} - B^{-1} &= A^{-1} - (L A)^{-1} \\ &= A^{-1} - A^{-1} L^{-1} \\ &= A^{-1} [I - L^{-1}] \\ &= A^{-1} [L - I] \geq \theta \end{aligned}$$

$$\text{et comme : } A^{-1} \geq \theta \quad \text{et} \quad B^{-1} \geq \theta \quad \text{donc } A^{-1} \geq B^{-1}.$$

II - THEOREMES DES DECOMPOSITIONS INCOMPLETES DES M-MATRICES

On désignera par P_n l'ensemble des paires d'indices (i, j) tel que $i \neq j$; P_n est appelé aussi ensemble des positions hors-diagonales dans A . $P_n = \{(i, j) / i \neq j ; i, j \in I\}$.

On dit que $M-N$ est un splitting de A si $A = M-N$ avec N une matrice non nulle.

On dit que A est incomplètement décomposable s'il existe L et U matrices respectivement triangulaire inférieure et triangulaire supérieure, et une matrice R non nulle telle que $A = LU - R$.

On dit qu'une matrice A admet une décomposition incomplète positive si A est incomplètement décomposable et si : $L_{ii} > 0$ et $U_{ii} > 0$ pour $i \in I$.

Les théorèmes énoncés dans ce paragraphe nous montrent que toute M -matrice admet des décompositions incomplètes positives et nous donnent un procédé pour le calcul effectif des matrices L , U et R .

Théorème 1.4 : (Meijerink - Van der Vorst)

Soit $A = (a_{ij})$ une M -matrice, alors pour tout $P \subseteq P_n$ il existe une matrice triangulaire inférieure $L = (L_{ij})$ de diagonale unité, une matrice triangulaire supérieure $U = (U_{ij})$ de diagonale positive et une matrice $R = (r_{ij})$ non négative telles que :

$$\begin{cases} L_{ij} = 0 & \text{si } (i, j) \in P \\ U_{ij} = 0 & \text{si } (i, j) \in P \\ r_{ij} = 0 & \text{si } (i, j) \notin P \end{cases}$$

et telle que $A = LU - R$.

Les facteurs L , U et R sont uniques.

Preuve :

Nous nous limiterons à rappeler le procédé utilisé par Meijerink - Van der Vorst [15] pour la démonstration de ce théorème, qui se fait par la construction des matrices L , U et R en $(n-1)$ étapes.

On considère les matrices suivantes :

$$(1.2) \quad \begin{cases} A^k = (a_{ij}^k) \\ \tilde{A}^k = (\tilde{a}_{ij}^k) \\ L^k = (L_{ij}^k) \\ R^k = (r_{ij}^k) \end{cases}$$

définies par la relation de récurrence suivante :

$$(1.3) \quad \begin{cases} A^0 = A \\ \text{et pour } k = 1, 2, \dots, n-1 \\ \tilde{A}^k = A^{k-1} + R^k \\ A^k = L^k \tilde{A}^k \end{cases}$$

avec :

$$(1.4) \quad \begin{cases} R^k = (r_{ij}^k) = \begin{cases} r_{kj}^k = -a_{kj}^{k-1} & \text{si } (k, j) \in P \\ r_{ik}^k = -a_{ik}^{k-1} & \text{si } (i, k) \in P \\ 0 & \text{sinon} \end{cases} \end{cases}$$

et L^k est la matrice de l'élimination de Gauss de la k° colonne à partir de la k° ligne de $\tilde{A}^k$:

$$(1.5) \quad L^k =$$

A la $(n-1)^{\circ}$ étape, on aura :

$$A^{n-1} = L^{n-1} \tilde{A}^{n-1}$$

et en tenant compte de la relation (1.3), on a :

$$A^{n-1} = [L^{n-1} L^{n-2} \dots L^1] [A - \sum_{k=1}^{n-1} R^k]$$

et en posant :

$$(1.6) \quad \begin{aligned} U &= A^{n-1} \\ L &= [L^{n-1} L^{n-2} \dots L^1]^{-1} \\ R &= \sum_{k=1}^{n-1} R^k \end{aligned}$$

on a :

$$A = LU - R.$$

Cas particulier : Cas où A est en plus symétrique.

Théorème 1.5 : (Meijerink - Van der Vorst)

Soit A une M -matrice symétrique ; alors pour tout $P \subseteq P_n$,

tel que si $(i, j) \in P \Rightarrow (j, i) \in P$, il existe une matrice unique triangulaire inférieure $L = (L_{ij})$ à diagonale positive et une matrice

$R = (r_{ij})$ non négative telles que :

$$L_{ij} = 0 \quad \text{si } (i, j) \in P$$

$$r_{ij} = 0 \quad \text{si } (i, j) \notin P$$

et
$$A = LL^T - R.$$

Remarques 1.1 :

1 - La décomposition incomplète de A pour $P = \emptyset$ ($\emptyset$: ensemble vide) correspond à la décomposition complète de A (car $R = 0$).

2 - Les splittings de Gauss - Seidel et de Jacobi font partie de ces décompositions incomplètes :

- Pour Gauss-Seidel : $P = \{(i, j) / i \neq j \text{ et } i < j ; i, j \in I\}$
- Pour Jacobi : $P = P_n = \{(i, j) / i \neq j ; i, j \in I\}$.

3 - Les splittings provenant des méthodes de résolution du système $Ax = b$ par bloc font aussi partie des décompositions incomplètes de la matrice A .

4 - Le calcul de la matrice R ne sera pas nécessaire puisqu'on peut remplacer R par $LU - A$.

III - THEOREMES DE CONVERGENCE ET DE STABILITE NUMERIQUE

On a à résoudre le système linéaire :

(1.7) $Ax = b$; avec A une matrice non singulière d'ordre n , x et b deux vecteurs de $\mathbb{R}^n$ par une méthode indirecte ; on décompose A en un splitting : $A = M - N$ et on considère la méthode itérative :

(1.8) $Mx_{i+1} = Nx_i + b$ pour $i \geq 0$ et x_0 un vecteur initial choisi de $\mathbb{R}^n$ et en remplaçant N par $M - A$ alors le système (1.8) sera équivalent à :

(1.9) $M(x_{i+1} - x_i) = b - Ax_i \quad i \geq 0 \quad \text{et} \quad x_0 \text{ donné de } \mathbb{R}^n$
et si on considère une décomposition incomplète de A :

$$A = LU - R$$

résoudre (1.8) serait équivalent à résoudre le système :

$$(1.10) \quad \begin{cases} U \Delta x_i = y_i & \text{système triangulaire supérieur} \\ L y_i = b - Ax_i & \text{système triangulaire inférieur} \\ x_{i+1} = \Delta x_i + x_i & \text{pour } i \geq 0 \text{ et } x_0 \text{ un vecteur choisi dans } \mathbb{R}^n \end{cases}$$

Définition 1.6 :

Soient A , M , N et $A = M - N$. On dit qu'on a un splitting régulier de la matrice A si M est non singulière, $M^{-1} \geq 0$ et $N \geq 0$.

Proposition 1.2 :

Soient A une M -matrice et $A = M - N$ un splitting régulier de A alors on a $\rho(M^{-1}N) = \frac{\rho(A^{-1}N)}{1 + \rho(A^{-1}N)} < 1$, et la méthode itérative (1.8) converge vers l'unique solution de (1.7) $\forall x_0$.

Preuve : cf. Varga [28] ou Meurant - Golub [17].

Corollaire 1.3 :

Soient A une M -matrice et deux splittings réguliers de A .
 $A = M_1 - N_1$ et $A = M_2 - N_2$ avec $N_2 \leq N_1$ (on dira que les deux splittings sont comparables). On a alors :

$$\rho(M_2^{-1}N_2) \leq \rho(M_1^{-1}N_1) < 1.$$

Preuve : cf. varga [28].

On va montrer que les décompositions incomplètes d'une M -matrice A sont des splittings réguliers de A .

Théorème 1.7 : (Meijerink - Van der Vorst)

Soit A une M -matrice, alors toute décomposition incomplète associée à $P \subseteq P_n$ est un splitting régulier de A et la méthode itérative (1.10) associée à cette décomposition converge vers l'unique solution de (1.7) pour tout x_0 .

Remarques 1.2 :

1 - Si $L_1 U_1 = R_1$ est une décomposition incomplète de la M -matrice A associée à $P_1 \subseteq P_n$, et $L_2 U_2 = R_2$ est une deuxième décomposition incomplète associée à P_2 et si $R_1 \geq R_2$ alors on a $P_2 \subseteq P_1$.

2 - D'après la remarque précédente, on constate que le meilleur choix de P , pour qu'on ait la meilleure convergence de la méthode itérative (1.5), est $P = \mathbb{I}$ qui correspond à la décomposition complète de A .

3 - Soit $P \subseteq P_n$, et si on considère la décomposition incomplète associée à P , définie par le théorème 1.4

$$A = \tilde{L} \tilde{U} - R$$

en utilisant la disposition de Crout, cf. Gastinel [9], pour $A+R = \tilde{L} \tilde{U}$, on a les formules de calcul suivantes :

$$(1.11) \quad \left\{ \begin{array}{ll} \tilde{L}_{ii} = 1 \text{ pour } i = 1, \dots, n & \\ \tilde{L}_{ij} = \begin{cases} (a_{ij} - \sum_{k=1}^{j-1} \tilde{L}_{ik} \tilde{U}_{kj}) / \tilde{U}_{jj} & \text{si } (i, j) \notin P \text{ et } i > j \\ 0 & \text{sinon} \end{cases} & \\ \tilde{U}_{ii} = a_{ii} - \sum_{k=1}^{i-1} \tilde{L}_{ik} \tilde{U}_{ki} & \text{pour } i \in I \\ \tilde{U}_{ij} = \begin{cases} a_{ij} - \sum_{k=1}^{i-1} \tilde{L}_{ik} \tilde{U}_{kj} & \text{si } (i, j) \notin P \text{ et } j > i \\ 0 & \text{sinon} \end{cases} & \end{array} \right.$$

et si on pose :

$$L = \tilde{L} \text{ et } U = \Sigma \tilde{U} \text{ avec } \Sigma = (\sigma_i) = (\tilde{U}_{ii}) : \text{matrice diagonale}$$

les formules (1.11) seront :

$$(1.12) \quad \left\{ \begin{array}{ll} \left. \begin{array}{l} \sigma_i = a_{ii} - \sum_{k=1}^{i-1} L_{ik} \sigma_k U_{ki} \\ L_{ii} = 1 \\ U_{ii} = 1 \end{array} \right\} & \text{pour } i \in I \\ L_{ij} = \begin{cases} \frac{1}{\sigma_j} [a_{ij} - \sum_{k=1}^{j-1} L_{ik} \sigma_k U_{kj}] & \text{si } (i, j) \notin P \text{ et } i > j \\ 0 & \text{sinon} \end{cases} & \\ U_{ij} = \begin{cases} \frac{1}{\sigma_i} [a_{ij} - \sum_{k=1}^{i-1} L_{ik} \sigma_k U_{kj}] & \text{si } (i, j) \notin P \text{ et } j > i \\ 0 & \text{sinon} \end{cases} & \end{array} \right.$$

4 - Si A est en plus symétrique, les formules (1.12) seront :

$$A = L \Sigma L^T - R$$

$$(1.13) \quad \left\{ \begin{array}{l} \sigma_i = a_{ii} - \sum_{k=1}^{i-1} L_{ik}^2 \sigma_k \quad \text{pour } i \in I \\ L_{ii} = 1 \quad \text{pour } i \in I \\ L_{ij} = \begin{cases} \frac{1}{\sigma_j} \left[a_{ij} - \sum_{k=1}^{j-1} L_{ik} \sigma_k L_{jk} \right] & \text{si } (i,j) \notin P \text{ et } i > j \\ 0 & \text{sinon} \end{cases} \end{array} \right.$$

La question qui se pose maintenant est de savoir si la méthode des décompositions incomplètes d'une M-matrice est stable ; pour cela, nous allons avoir besoin du lemme suivant :

Lemme 1.1 :

Soient $A = (a_{ij})$ une M-matrice et $B = (b_{ij})$ la matrice telles que :
 $a_{ij} \leq b_{ij} \leq 0$ pour $i \neq j$ et $a_{ii} \leq b_{ii}$; si $A = (a_{ij}^1)$ et $B^1 = (b_{ij}^1)$

les matrices obtenues respectivement à partir de A et B par l'élimination de Gauss de la première colonne à partir de la première ligne, alors on a :

$$(1.14) \quad \left\{ \begin{array}{l} a_{ij}^1 \leq b_{ij}^1 \leq 0 \quad \text{pour } i \neq j \\ 0 < a_{ii}^1 \leq b_{ii}^1 \end{array} \right.$$

et B^1 est une M-matrice.

Preuve :

La matrice B est une M-matrice car elle vérifie les hypothèses du théorème 1.2 et d'après le théorème 1.3, A^1 et B^1 sont des M-matrices. En appliquant les hypothèses du théorème aux formules de l'élimination de Gauss données par (1.1), on a :

$$a_{ij}^1 = a_{ij} - \frac{a_{i1} a_{1j}}{a_{11}} \leq b_{ij} - \frac{b_{i1} b_{1j}}{b_{11}} = b_{ij}^1 \leq 0 \quad \text{pour } i \neq j$$

$$\text{et} \quad 0 < a_{ii} - a_{i1} \frac{a_{1i}}{a_{11}} = a_{ii}^1 \leq b_{ii} - b_{i1} \frac{b_{1i}}{b_{11}} = b_{ii}^1$$

d'où la démonstration du théorème.

Théorème 1.8 :

Si $A = (a_{ij})$ est une M-matrice alors la méthode des décompositions incomplètes de A est au moins aussi stable numériquement que la méthode de décomposition complète (décomposition de Gauss) sans pivot.

Preuve :

Soient $A = LU - R$ la décomposition incomplète de la M-matrice A associée à $P \subseteq P_n$ et définie par le théorème 1.4

$A = \tilde{L}\tilde{U}$ la décomposition complète de A
(simple élimination de Gauss).

Il s'agit alors de montrer que $U_{ii} \geq \tilde{U}_{ii}$ pour $i \in I$ ($\tilde{U}_{ii} > 0$ théorème 1.1).

Pour cela, on se sert des formules (1.3) - (1.5) utilisées dans le procédé pour le calcul de L, U et R ; ainsi la démonstration se fera étape par étape : $k = 1$ on introduit des zéros dans des positions hors-diagonales de

A : $\tilde{A}_1 = A + R_1$ d'où si $\tilde{A}_1 = (\tilde{a}_{ij}^1)$: $\tilde{a}_{ii}^1 = a_{ii} > 0$ et $0 \geq a_{ij}^1 \geq a_{ij}$ pour $i \neq j$

donc A et $\tilde{A}_1$ vérifient les hypothèses du lemme 1.1 donc les matrices

obtenues de A et $\tilde{A}_1$ par l'élimination de Gauss de la première colonne

à partir de la première ligne vérifient les formules (1.14) qui sont aussi les hypothèses du lemme ; et on recommence le même raisonnement pour

$k = 1, \dots, n-1$, à l'étape $k = n-1$ on trouve que $U_{ii} \geq \tilde{U}_{ii} > 0$
pour tout $i \in I$.

- CHAPITRE 2 -
=====

DECOMPOSITIONS INCOMPLETES DES H-MATRICES
--

Introduction :

Dans ce chapitre, nous rappelons, tout d'abord, quelques propriétés des H-matrices. Ces propriétés sont utilisées, ensuite, pour obtenir la décomposition incomplète de toute H-matrice (symétrique ou non). On énoncera de plus un résultat assurant la convergence de la méthode de point fixe associée à une telle décomposition incomplète de la H-matrice.

On énoncera ensuite quelques nouvelles caractérisations des H-matrices qui ont un rapport avec les décompositions incomplètes de ces dernières.

En pratique, ces décompositions incomplètes se combinent avec des méthodes d'accélération de la convergence, que l'on étudiera dans le chapitre 4 : la méthode du gradient conjugué généralisée [7], la méthode du double-gradient conjugué [32] et la méthode d'orthogonalisation-minimisation [30].

- CHAPITRE 2 -

1 - DEFINITION ET PROPRIETES DES H-MATRICES

Soit $A = (a_{ij})$ une matrice carrée d'ordre n à coefficients réels
 $A = (a_{ij}) = D - E - F$
 $N(A) = |D| - |E| - |F|$ la Z-matrice associée à A .

Définition 2.1 :

On dit que $A = (a_{ij})$ est une H-matrice si $N(A)$ est une M-matrice.

Proposition 2.1 :

Si A est une H-matrice, alors :

- 1 - A est régulière
- 2 - $|A^{-1}| \leq [N(A)]^{-1}$.

Preuve : cf. Robert [23].

Proposition 2.2 :

Soit A une matrice, alors des conditions nécessaires et suffisantes pour que A soit une H-matrice sont :

- 1 - Les éléments diagonaux de A sont non nuls
- 2 - $\rho(|J_A|) = \rho(J_{N(A)}) < 1$.

Preuve : cf. R. Varga [27]

Théorème 2.1 :

Soient $A = (a_{ij})$ une H-matrice et $B = (b_{ij})$ la matrice telles que :

$$|a_{ii}| \leq |b_{ii}| \quad \text{et} \quad -|a_{ij}| \leq -|b_{ij}| \quad \text{pour } i \neq j ;$$

alors B est aussi une H-matrice et on a :

$$[N(A)]^{-1} \geq [N(B)]^{-1} \geq |B^{-1}|.$$

Preuve :

Si on note $A = (a_{ij}) = D_A - E_A - F_A$ et $B = (b_{ij}) = D_B - E_B - F_B$,

on a : $N(A) = |D_A| - |E_A| - |F_A|$ et $N(B) = |D_B| - |E_B| - |F_B|$

et les hypothèses faites dans le théorème s'expriment matriciellement par :

$$|D_A| \leq |D_B|, \quad |E_B| \leq |E_A| \quad \text{et} \quad |F_B| \leq |F_A|.$$

Or ceci vérifie les hypothèses du théorème 1.2 d'où

$$[N(A)]^{-1} \geq [N(B)]^{-1} \text{ et } B \text{ est une H-matrice}$$

et tenant compte de la proposition 2.1, on a :

$$[N(A)]^{-1} \geq [N(B)]^{-1} \geq |B^{-1}|.$$

De ce théorème, on a le corollaire suivant :

Corollaire 2.1 :

Soient $A = (a_{ij})$ une H-matrice et $C = (c_{ij})$ la matrice obtenue à partir de A en introduisant des zéros dans des positions hors-diagonales dans A, alors C est aussi une H-matrice et on a :

$$[N(A)]^{-1} \geq [N(C)]^{-1} \geq |C^{-1}|.$$

On peut généraliser le théorème 2.1 : on considère

$$A \in M_{n,n}(\mathbb{C}) \text{ et } \Omega^d(A) = \{B \in M_{n,n}(\mathbb{C}) / |a_{ii}| \leq |b_{ii}|$$

$$\text{et } |b_{ij}| \leq |a_{ij}| \quad \text{pour } i \neq j\}.$$

Théorème 2,2 :

Si A est une H -matrice alors $\forall B \in \Omega^d(A)$, B est aussi une H -matrice et on a $[N(A)]^{-1} \geq [N(B)]^{-1} \geq |B^{-1}|$.

Preuve :

On note $A = (a_{ij}) = D_A - E_A - F_A$ et $B = (b_{ij}) = D_B - E_B - F_B$

d'où $N(A) = |D_A| - |E_A| - |F_A|$ et $N(B) = |D_B| - |E_B| - |F_B|$

et $J_A = D_A^{-1}[E_A + F_A]$ et $J_B = D_B^{-1}[E_B + F_B]$

d'où $|J_A| = |D_A^{-1}[E_A + F_A]| = |D_A^{-1}|[|E_A| + |F_A|]$

$$= |D_A|^{-1}[|E_A| + |F_A|] = J_{N(A)}$$

et $|J_B| = |D_B^{-1}[E_B + F_B]| = |D_B|^{-1}[|E_B| + |F_B|] = J_{N(B)}$

or par hypothèse $B \in \Omega^d(A)$ donc

$$\frac{|b_{ij}|}{|b_{ii}|} \leq \frac{|a_{ij}|}{|a_{ii}|}$$

soit matriciellement : $|J_B| \leq |J_A|$

et comme A est une H -matrice alors

$$\rho(|J_B|) \leq \rho(|J_A|) < 1$$

donc B est une H -matrice et on a :

$$[N(B)]^{-1} \geq |B^{-1}|.$$

Montrons que : $[N(A)]^{-1} \geq [N(B)]^{-1}$.

On a : $|J_A| = I - |D_A|^{-1}N(A)$ et $|J_B| = I - |D_B|^{-1}N(B)$

d'où : $N(A) = |D_A|[I - |J_A|]$ et $N(B) = |D_B|[I - |J_B|]$

or : $|J_B| \leq |J_A|$ et $\rho(|J_B|) \leq \rho(|J_A|) < 1$

donc : $(I - |J_A|)^{-1}$ et $(I - |J_B|)^{-1}$ existent et sont non négatives

et on a : $(I - |J_A|)^{-1} \geq (I - |J_B|)^{-1} \geq \theta$

et par hypothèse : $|D_A|^{-1} \geq |D_B|^{-1} > \theta$

donc : $[N(A)]^{-1} = [I - |J_A|]^{-1} |D_A|^{-1} \geq [I - |J_B|]^{-1} |D_B|^{-1} = [N(B)]^{-1}.$

Théorème 2.3 :

Soient $A = (a_{ij})$ une H-matrice et $B = (b_{ij})$ la M-matrice telles que (en particulier $B = N(A)$) :

$$(2.1) \quad \begin{cases} b_{ii} \leq |a_{ii}| \\ |a_{ij}| \leq -b_{ij} \quad \text{pour } i \neq j \end{cases}$$

si $\tilde{A} = (\tilde{a}_{ij})$ (resp. $\tilde{B} = (\tilde{b}_{ij})$) est la matrice obtenue à partir de A (resp. B) par l'élimination de Gauss de la k° colonne à partir de la k° ligne ; alors $\tilde{A}$ est une H-matrice et les matrices $\tilde{A}$ et $\tilde{B}$ vérifient encore les formules (2.1) et on a :

$$[N(\tilde{A})]^{-1} \leq \tilde{B}^{-1}.$$

Preuve :

On sait que l'élimination de Gauss, sur une matrice A , de la k° colonne à partir de la k° ligne ($k = 1, \dots, n-1$) laisse les k premières lignes de A inchangées. Les formules de l'élimination de Gauss sur A et B sont :

Pour $\tilde{A} = (\tilde{a}_{ij})$

$$(2.2) \quad \begin{cases} \tilde{a}_{ij} = a_{ij} & \text{pour } i = 1, \dots, k \quad \text{et } j \in I \\ \tilde{a}_{ij} = a_{ij} - \frac{a_{ik} a_{kj}}{a_{kk}} & \text{pour } i = k+1, \dots, n \quad \text{et } j \in I \end{cases}$$

Pour $\tilde{B} = (\tilde{b}_{ij})$

$$(2.3) \quad \begin{cases} \tilde{b}_{ij} = b_{ij} & \text{pour } i = 1, \dots, k \quad \text{et } j \in I \\ \tilde{b}_{ij} = b_{ij} - \frac{b_{ik} b_{kj}}{b_{kk}} & \text{pour } i = k+1, \dots, n \quad \text{et } j \in I. \end{cases}$$

Pour démontrer le théorème, on va comparer les matrices $N(\tilde{A})$ la Z-matrice associée à $\tilde{A}$ et la matrice $\tilde{B}$, en deux étapes.
Compte tenu du théorème 1.3, la matrice $\tilde{B}$ est une M-matrice, donc :

$$\begin{cases} \tilde{b}_{ii} > 0 & \text{pour } i \in I \\ \tilde{b}_{ij} \leq 0 & \text{pour } i \neq j \quad \text{et } i, j \in I. \end{cases}$$

La matrice $N(\tilde{A})$ est donnée par les formules suivantes :

$$(2.4) \quad \begin{cases} |\tilde{a}_{ii}| = |a_{ii}| & \text{pour } i = 1, \dots, k \\ |\tilde{a}_{ii}| = |a_{ii} - \frac{a_{ik} a_{ki}}{a_{kk}}| & \text{pour } i = k+1, \dots, n \\ -|\tilde{a}_{ij}| = -|a_{ij}| & \text{pour } i \neq j \quad i = 1, \dots, k \quad \text{et } j \in I \\ -|\tilde{a}_{ij}| = -|a_{ij} - \frac{a_{ik} a_{kj}}{a_{kk}}| & \text{pour } i \neq j \quad i = k+1, \dots, n \quad \text{et } j \in I. \end{cases}$$

1ère étape :

$B = N(A)$ est la M-matrice associée à A .

Les formules (2.3) deviennent alors :

$$(2.5) \quad \begin{cases} \tilde{b}_{ii} = |a_{ii}| & \text{pour } i = 1, \dots, k \\ \tilde{b}_{ii} = |a_{ii}| - \left| \frac{a_{ik} a_{ki}}{a_{kk}} \right| & \text{pour } i = k+1, \dots, n \\ \tilde{b}_{ij} = -|a_{ij}| & \text{pour } i \neq j \quad i = 1, \dots, k \quad \text{et } j \in I \\ \tilde{b}_{ij} = -|a_{ij}| - \left| \frac{a_{ik} a_{kj}}{a_{kk}} \right| & \text{pour } i \neq j \quad i = k+1, \dots, n \quad \text{et } j \in I \end{cases}$$

et on va comparer les éléments donnés par les formules (2.4) et (2.5).

Pour $i = k+1, \dots, n$ on a :

$$|\tilde{a}_{ii}| = |a_{ii} - \frac{a_{ik} a_{ki}}{a_{kk}}| \geq ||a_{ii}| - |\frac{a_{ik} a_{ki}}{a_{kk}}|| = |a_{ii}| - |\frac{a_{ik} a_{ki}}{a_{kk}}|$$

$$= \tilde{b}_{ii} > 0.$$

Pour $i \neq j$ et $i = k+1, \dots, n, j \in I$ on a :

$$-|\tilde{a}_{ij}| = -|a_{ij} - \frac{a_{ik} a_{kj}}{a_{kk}}| \geq -|a_{ij}| - |\frac{a_{ik} a_{kj}}{a_{kk}}| = \tilde{b}_{ij},$$

donc :

$$\begin{cases} |\tilde{a}_{ii}| \geq \tilde{b}_{ii} & \text{pour } i \in I \\ -|\tilde{a}_{ij}| \geq \tilde{b}_{ij} & \text{pour } i \neq j \text{ et } i, j \in I \end{cases}$$

et les matrices $\tilde{B}$ et $N(\tilde{A})$ vérifient les hypothèses du théorème 1.2
donc $N(\tilde{A})$ est une M-matrice et on a :

$$[N(\tilde{A})]^{-1} \leq \tilde{B}^{-1}.$$

2ème étape :

B et A vérifient les formules (2.1) ; on va comparer les matrices $\tilde{B}$ et $N(\tilde{A})$ qui sont des M-matrices.

Pour $i = k+1, \dots, n$ on a :

$$|\tilde{a}_{ii}| = |a_{ii} - \frac{a_{ik} a_{ki}}{a_{kk}}| \geq |a_{ii}| - |\frac{a_{ik} a_{ki}}{a_{kk}}| > 0$$

$$\tilde{b}_{ii} = b_{ii} - \frac{b_{ik} b_{ki}}{b_{kk}}$$

d'où :

$$|a_{ii}| - |\frac{a_{ik} a_{ki}}{a_{kk}}| - \tilde{b}_{ii} = (|a_{ii}| - b_{ii}) + (\frac{b_{ik} b_{ki}}{b_{kk}} - |\frac{a_{ik} a_{ki}}{a_{kk}}|) \geq 0$$

car on a, par hypothèses :

$$(a_{ii} - b_{ii}) \geq 0 \quad \text{et} \quad -|a_{ij}| \geq b_{ij} \quad \text{pour } i \neq j$$

d'où :

$$\frac{b_{ik} b_{ki}}{b_{kk}} - |\frac{a_{ik} a_{ki}}{a_{kk}}| \geq 0$$

donc : $|\tilde{a}_{ij}| \geq \tilde{b}_{ij}$.

Pour $i \neq j$ et $i = k+1, \dots, n$, $j \in I$ on a :

$$-|\tilde{a}_{ij}| = -|a_{ij} - \frac{a_{ik} a_{kj}}{a_{kk}}| \geq -|a_{ij}| - |\frac{a_{ik} a_{kj}}{a_{kk}}|$$

d'où :

$$-|a_{ij}| - |\frac{a_{ik} a_{kj}}{a_{kk}}| - \tilde{b}_{ij} = (-|a_{ij}| - \tilde{b}_{ij}) + (\frac{b_{ik} b_{kj}}{b_{kk}} - |\frac{a_{ik} a_{kj}}{a_{kk}}|) \geq 0$$

donc : $|\tilde{a}_{ij}| \geq \tilde{b}_{ij}$

et on a :

$$\begin{cases} |\tilde{a}_{ii}| \geq \tilde{b}_{ii} & \text{pour } i \in I \\ -|\tilde{a}_{ij}| \geq \tilde{b}_{ij} & \text{pour } i \neq j \text{ et } i, j \in I \end{cases}$$

donc les matrices $\tilde{B}$ et $\tilde{A}$ vérifient les formules (2.1) qui sont les hypothèses du théorème, et $\tilde{B}$ et $N(\tilde{A})$ vérifient les hypothèses du théorème 1.2 donc on a $[N(A)]^{-1} \leq \tilde{B}^{-1}$, et $\tilde{A}$ est une H-matrice.

Théorème 2.4 :

Si $A = (a_{ij})$ est une H-matrice alors :

1 - Toutes les matrices mineures principales A_k ($k = 1, \dots, n$) de A sont des H-matrices.

2 - Tous les mineurs principaux de A sont non nuls et on a :

$$|\det(A_k)| \geq \det[N(A_k)].$$

3 - Il existe L et U matrices triangulaires respectivement inférieure et supérieure à éléments diagonaux non nuls telles que $A = LU$.

Preuve :

A est une H-matrice donc la Z-matrice associée à A est une M-matrice.

1 - Compte tenu du théorème 1.1 et du corollaire 1.1 on a $N(A_k)$

pour $k = 1, \dots, n$ sont des M-matrices et leurs déterminants sont positifs,

donc les matrices mineures principales, A_k pour $k \in I$, de A sont des H-matrices et leurs déterminants sont non nuls,

2 - Montrons que : $|\det(A_k)| \geq \det[N(A_k)]$ pour $k \in I$. Pour cela, on va se servir du théorème 2.3, soit $k \in I$ notons $I_k = \{1, 2, \dots, k\} \subseteq I$ on a : $A_k = (a_{ij})$ pour $i, j \in I_k$ est une H-matrice.

Notons $B_k = N(A_k) = (b_{ij})$ pour $i, j \in I_k$ la M-matrice associée à A_k .

On a : $b_{ii} = |a_{ii}|$ et $-b_{ij} = |a_{ij}|$ pour $i \neq j$ et $i, j \in I_k$

donc A_k et B_k vérifient les hypothèses du théorème 2.3, et si on pose :

$\tilde{A}_k = (\tilde{a}_{ij})$ et $\tilde{B}_k = (\tilde{b}_{ij})$ les matrices obtenues respectivement de A_k et

B_k par l'élimination de Gauss de la première colonne à partir de la pre-

mière ligne on trouve que $\tilde{A}_k$ et $\tilde{B}_k$ vérifient les formules (2.1) donc elles

vérifient les hypothèses du théorème 2.3 et on applique ainsi les élimina-

tions de Gauss de la seconde colonne à partir de la seconde ligne de $\tilde{A}_k$

et $\tilde{B}_k$ et on retrouve les formules (2.1). Ainsi à l'étape $(k-1)$ on aura,

si on note :

$A_k^{k-1} = (a_{ij}^{k-1})$ la matrice obtenue à partir de A_k par l'élimination de Gauss

$B_k^{k-1} = (b_{ij}^{k-1})$ la matrice obtenue à partir de B_k par l'élimination de Gauss

L et $\tilde{L}$ les matrices de l'élimination de Gauss sur A_k et B_k

on a alors et compte tenu du théorème 2.5 et des formules (2.1) :

$$-|a_{ij}^{k-1}| \geq b_{ij}^{k-1} \text{ pour } i \neq j \text{ et } i, j \in I_k$$

$$\text{et } |a_{ii}^{k-1}| \geq b_{ii}^{k-1} \text{ pour } i \in I_k$$

et $A_k^{k-1} = L A_k$ une matrice triangulaire supérieure

$$B_k^{k-1} = \tilde{L} B_k \text{ une matrice triangulaire supérieure}$$

$L = (L_{ij})$ est une matrice triangulaire inférieure à diagonale unitée

$\tilde{L} = (\tilde{L}_{ij})$ est une matrice triangulaire inférieure à diagonale unitée

d'où :

$$\det(A_k) \det(L) = \det(A_k^{k-1}) = \prod_{i=1}^k a_{ii}^{k-1}$$

or $\det(L) = \det(\tilde{L}) = 1$

d'où : $\det(A_k) = \prod_{i=1}^k [a_{ii}^{k-1}]$

$$\det(B_k) = \prod_{i=1}^k b_{ii}^{k-1}$$

et compte tenu du fait que : $|a_{ii}^{k-1}| \geq b_{ii}^{k-1}$

on a : $|\det(A_k)| = \prod_{i=1}^k |a_{ii}^{k-1}| \geq \prod_{i=1}^k b_{ii}^{k-1} = \det(B_k)$

et cela pour $k \in I$.

3 - La démonstration du point 2 est vraie pour $k = n$ c'est-à-dire pour $A_n = A$ donc l'existence des matrices L et U est assurée si on pose

$A^{n-1} = U$ et on a $L_{ii} = 1$ et $|U_{ii}| = |a_{ii}^{n-1}| \geq b_{ii}^{n-1} > 0$.

II - THEOREMES DES DECOMPOSITIONS INCOMPLETES DES H-MATRICES

On désignera par P_n l'ensemble des positions hors-diagonales dans une matrice A $P_n = \{(i, j) / i \neq j \text{ et } i, j \in I\}$.

On dira que la matrice A est incomplètement décomposable s'il existe L et U matrices respectivement triangulaire inférieure et triangulaire supérieure, et une matrice R non nulle telle que $A = LU - R$.

On dira que cette décomposition incomplète est régulière (*) si en plus : les éléments diagonaux de L et U sont non nuls.

Les théorèmes énoncés dans ce paragraphe nous montrent que toute H -matrice admet des décompositions incomplètes régulières.

Théorème 2.5 :

Soit $A = (a_{ij})$ une H -matrice carrée $n \times n$; alors pour tout $P \subseteq P_n$ il existe une matrice triangulaire inférieure $L = (L_{ij})$ de diagonale unité ($L_{ii} = 1$), une matrice triangulaire supérieure $U = (U_{ij})$ et une matrice $R = (r_{ij})$ telles que :

$$\begin{cases} L_{ij} = 0 & \text{si } (i, j) \in P \\ U_{ij} = 0 & \text{si } (i, j) \in P \\ r_{ij} = 0 & \text{si } (i, j) \notin P \end{cases}$$

et telle que $A = LU - R$ soit une décomposition incomplète régulière. Les facteurs L , U et R sont uniques.

Preuve :

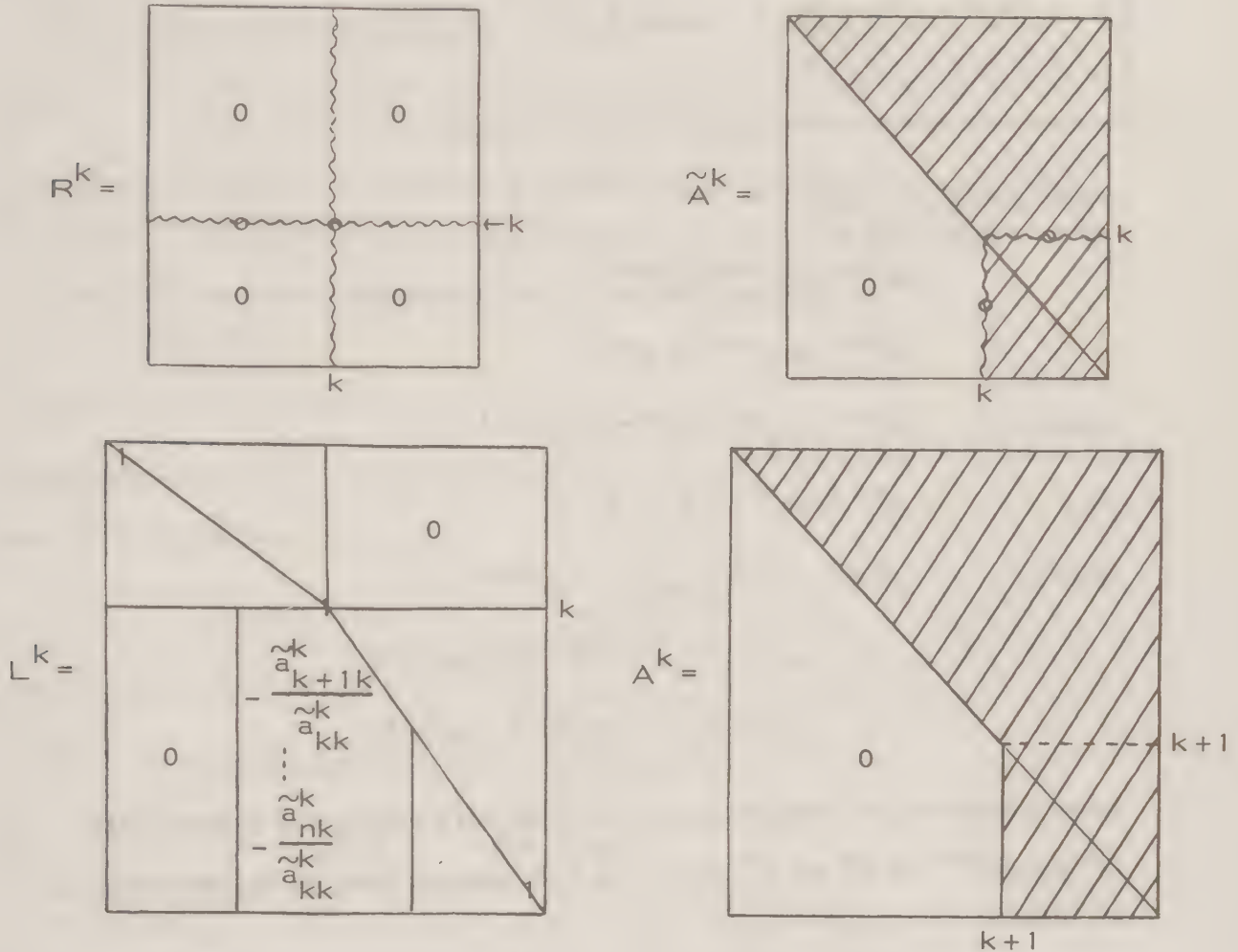
Nous allons proposer deux démonstrations de ce théorème.

1ère démonstration : qui reprend le procédé utilisé par Meijerink - Van der Vorst [15] pour les décompositions incomplètes des M -matrices. Cette démonstration se fait par la construction des matrices L , U et

(*) : Ne pas confondre avec un splitting régulier cf. définition 1.6.

R en $(n-1)$ étapes ; soient les matrices $\tilde{A}^k$, A^k , R^k et L^k définies par les formules (1.2) - (1.5).

Structure de ces matrices :



$A^0 = A$ est une H-matrice par hypothèse

$\tilde{A}^1 = A^0 + R^1$ est obtenue en introduisant des zéros dans des positions

hors-diagonales de A , donc, compte tenu du corollaire 2.1, $\tilde{A}^1$ est une

H-matrice $A^1 = L^1 \tilde{A}^1$ est obtenue par l'élimination de Gauss sur $\tilde{A}^1$ de

la première colonne à partir de la première ligne et compte tenu du théo-

rème 2.3, A^1 est une H-matrice.

Ainsi, pour $k = 1, \dots, n-1$, on a $\tilde{A}^k$ et A^k sont des H-matrices.

Remarques 2.1 :

1 - On a pour $m > k$: $L^k R^m = R^m$; ceci provient de la construction et de la structure des matrices L^k et R^m .

2 - $L^k L^m = L^k + L^m - I$ pour $k < m$

et $[L^k]^{-1} = 2I - L^k$.

Pour $k = n-1$, on a :

$$A^{n-1} = L^{n-1} \tilde{A}^{n-1}$$

$$\tilde{A}^{n-1} = A^{n-2} + R^{n-1}$$

d'où : $A^{n-1} = L^{n-1} [A^{n-2} + R^{n-1}]$

et $A^{n-2} = L^{n-2} \tilde{A}^{n-2}$

d'où :

$$\begin{aligned} A^{n-1} &= L^{n-1} [A^{n-2} + R^{n-1}] \\ &= L^{n-1} [L^{n-2} \tilde{A}^{n-2} + R^{n-1}] \\ &= L^{n-1} L^{n-2} [\tilde{A}^{n-2} + R^{n-1}] \end{aligned}$$

compte tenu de la remarque 2.1 et ainsi on remplace à chaque fois

$\tilde{A}^k$ par $A^{k-1} + R^k$ et A^k par $L^k \tilde{A}^k$ et compte tenu de la remarque 2.1 on obtient :

$$A^{n-1} = [L^{n-1} L^{n-2} \dots L^1] [A^0 + R^1 + \dots + R^{n-1}].$$

On pose alors :

$$U = A^{n-1}$$

$$L = [L^{n-1} L^{n-2} \dots L^1]^{-1} \text{ qui existe car } \det(L) = 1$$

$$R = \sum_{k=1}^{n-1} R^k.$$

On trouve alors :

$$A = LU - R$$

Or $U = A^{n-1}$ est une H-matrice triangulaire supérieure

donc $\det(U) = \det(A^{n-1}) = \prod_{i=1}^n a_{ii}^{n-1} \neq 0$

d'où : $U_{ii} = a_{ii}^{n-1} \neq 0$.

Compte tenu de la remarque 2.1, on a :

$$(2.6) \quad L = [L^{n-1} L^{n-2} \dots L^1]^{-1} = nI - \sum_{k=1}^{n-1} L^k$$

et L est une matrice triangulaire inférieure de diagonale unitée. Ainsi on a montré que pour tout $P \subseteq P_n$ il existe L , U et R telle que

$A = LU - R$ soit une décomposition incomplète régulière.

On va montrer que :

$$1 - U_{ij} = 0 \quad \text{si } (i, j) \in P.$$

On a posé $U = A^{n-1}$ d'où $U_{ij} = a_{ij}^{n-1}$

pour $k = 1$, on a :

$$\tilde{A}^1 = A + R^1 \quad \text{et} \quad A^1 = L^1 \tilde{A}^1$$

donc si $(1, j) \in P$ on aura $r_{1j}^1 = -a_{1j}$ pour $j > 1$

et $\tilde{a}_{1j}^1 = a_{1j}^1 = 0$. De même, à la k° étape, on a :

$$\tilde{a}_{kj}^k = a_{kj}^k = 0 \quad \text{si } (k, j) \in P \quad \text{pour } j > k$$

car $r_{kj}^k = -a_{kj}^{k-1}$ si $(k, j) \in P$ et $\tilde{a}_{kj}^k = a_{kj}^{k-1} + r_{kj}^k$.

Ainsi donc pour A^{n-1} qui est obtenue par l'élimination de Gauss à partir de $\tilde{A}^{n-1}$ sans changer les $(n-1)$ premières lignes

donc $U_{ij} = 0$ si $(i, j) \in P$.

$$2 - L_{ij} = 0 \quad \text{si } (i, j) \in P.$$

Par construction des L^k on a :

$\tilde{A}^k = A^{k-1} + R^k$ et la colonne de L^k est donnée par :

$$-\frac{\tilde{a}_{jk}^k}{\tilde{a}_{kk}^k} = -\frac{a_{jk}^{k-1}}{\tilde{a}_{kk}^k} - \frac{r_{jk}^k}{\tilde{a}_{kk}^k} \quad \text{pour } j = k+1, \dots, n$$

et si $(j, k) \in P$ on a $r_{jk}^k = -a_{jk}^{k-1}$

donc
$$-\frac{a_{jk}^k}{a_{kk}^k} = 0 = L_{jk}^k$$

or comme :

$$L = nI - \sum_{k=1}^{n-1} L^k$$

donc $L_{ij} = 0$ si $(i, j) \in P$.

3 - $r_{ij} = 0$ si $(i, j) \notin P$.

On a par construction des R^k :

$$r_{ij}^k = 0 \quad \text{si } (i, j) \notin P$$

et comme $R = \sum_{k=1}^{n-1} R^k$ alors :

$$r_{ij} = 0 \quad \text{si } (i, j) \notin P.$$

4 - L , U et R sont uniques.

On remarque qu'une fois qu'on se donne $P \subseteq P_n$ on se donne alors R par la construction utilisée. D'où, si :

$$A = LU - R = L'U' - R$$

ceci donne que $LU = L'U'$

ou bien $L^{-1}L' = U'U^{-1}$

or $L^{-1}L'$ est le produit de deux matrices triangulaires inférieures à diagonales unitées et $U'U^{-1}$ est le produit de deux matrices triangulaires supérieures donc

$$L^{-1}L' = U'U^{-1} = I$$

et par conséquent : $L' = L$ et $U' = U$.

2ème démonstration : cette démonstration repose sur la remarque suivante :

une matrice A est incomplètement décomposable si :

1 - La matrice A est complètement décomposable (élimination de Gauss). Ceci est vrai si tous ses mineurs principaux sont non nuls.

2 - Chaque fois qu'on introduit des zéros dans des positions hors-diagonales dans A, les mineurs principaux de la nouvelle matrice ainsi obtenue sont non nuls.

Or pour une H-matrice A, le corollaire 2.1 garantit que la matrice obtenue à partir de A en introduisant des zéros dans des positions hors-diagonales reste une H-matrice et le théorème 2.4 nous montre que les mineurs principaux d'une H-matrice sont non nuls, donc le corollaire 2.1 et le théorème 2.4 assurent le deuxième point de la remarque ci-haut.

De même le théorème 2.3 et le théorème 2.4 assurent le premier point ; donc toute H-matrice admet une décomposition incomplète régulière pour tout $P \subseteq P_n$.

Exemple 2.1 :

$$A = \begin{bmatrix} -3 & 1 & -1 \\ 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} \quad N(A) = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix} \quad \text{et} \quad [N(A)]^{-1} = \frac{1}{2} \begin{bmatrix} 2 & 3 & 1 \\ 2 & 5 & 1 \\ 2 & 4 & 2 \end{bmatrix}$$

A est une H-matrice car $[N(A)]^{-1} \geq 0$.

On choisit $P = \{(1, 2), (2, 3)\}$

et on utilise la construction de L, U et R définie par les formules (1.2) - (1.5)

$$R^1 = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{d'où} \quad \tilde{A}^1 = A + R^1 = \begin{bmatrix} -3 & 0 & -1 \\ 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix}$$

$$L^1 = \begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 1/3 & 0 & 1 \end{bmatrix} \quad \text{d'où} \quad A^1 = L^1 \tilde{A}^1 = \begin{bmatrix} -3 & 0 & -1 \\ 0 & 1 & -1/3 \\ 0 & -1 & 5/3 \end{bmatrix}$$

$$R^2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1/3 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{et} \quad \tilde{A}^2 = A^1 + R^2 = \begin{bmatrix} -3 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & -1 & 5/3 \end{bmatrix}$$

$$L^2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad \text{d'où} \quad A^2 = L^2 \tilde{A}^2 = \begin{bmatrix} -3 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 5/3 \end{bmatrix}$$

d'où :

$$R = R^1 + R^2 = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 1/3 \\ 0 & 0 & 0 \end{bmatrix} \quad U = A^2 = \begin{bmatrix} -3 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 5/3 \end{bmatrix}$$

$$L = [L^2 L^1]^{-1} = 3I - L^1 - L^2 = \begin{bmatrix} 1 & 0 & 0 \\ -1/3 & 1 & 0 \\ -1/3 & -1 & 1 \end{bmatrix}$$

et on a : $A = LU - R$.

Remarques 2.2 :

1 - Les splittings de Jacobi et de Gauss-Seidel d'une H-matrice font partie des décompositions incomplètes de cette matrice :

- Pour Jacobi : $P = P_n = \{(i, j) / i \neq j\}$

- Pour Gauss-Seidel : $P = \{(i, j) / i < j\}$.

2 - Les méthodes de résolutions d'un système linéaire par bloc font aussi partie des décompositions incomplètes.

3 - Les formules (1.11) et (1.12) exprimées dans les remarques 1.2 sont aussi vraies pour une H-matrice.

Cas particuliers : Cas où A est une H -matrice symétrique.

Théorème 2.6 :

Soit A une H -matrice symétrique ; alors pour tout $P \subseteq P_n$ tel que si $(i, j) \in P$ implique que $(j, i) \in P$, il existe une matrice triangulaire inférieure $L = (L_{ij})$ à diagonale unité, une matrice diagonale $\Sigma = (\sigma_i)$ et une matrice $R = (r_{ij})$ telles que :

$$\begin{cases} L_{ij} = 0 & \text{si } (i, j) \in P \\ r_{ij} = 0 & \text{si } (i, j) \notin P \end{cases}$$

et telles que $A = L \Sigma L^T - R$ soit une décomposition incomplète régulière $(\sigma_i \neq 0)$ de A . Les facteurs L , Σ et R sont uniques.

Preuve :

Le théorème 2.5 nous garantit l'existence et l'unicité de $\tilde{L}$, $\tilde{U}$ et R telles que $A = \tilde{L} \tilde{U} - R$ et si on pose $\tilde{L} = L$ et $\tilde{U} = \Sigma U$ avec $\Sigma = (\sigma_i) = (\tilde{U}_{ii})$ et compte tenu de la symétrie de A et la symétrie de la décomposition alors on a $L^T = U$ d'où $A = L \Sigma L^T - R$ or $\tilde{U}_{ii} = \sigma_i \neq 0$ d'après le théorème 2.5.

Cas où A est une H -matrice à éléments diagonaux positifs.

Théorème 2.7 :

Soit A une H -matrice à éléments diagonaux positifs ; alors pour tout $P \subseteq P_n$ il existe une matrice triangulaire inférieure $L = (L_{ij})$ à diagonale unité, une matrice triangulaire supérieure $U = (U_{ij})$ à diagonale positive et une matrice $R = (r_{ij})$ telles que :

$$\begin{cases} L_{ij} = 0 & \text{si } (i, j) \in P \\ U_{ij} = 0 & \text{si } (i, j) \in P \\ r_{ij} = 0 & \text{si } (i, j) \notin P \end{cases}$$

et telles que $A = L U - R$.

III - THEOREMES DE CONVERGENCE ET DE STABILITE NUMERIQUE

On a à résoudre le système (1.7) avec A une H -matrice, par une méthode indirecte de la forme (1.10) en utilisant une décomposition incomplète de A . On va montrer que la méthode (1.10) converge pour toute décomposition incomplète de la H -matrice A .

Définition 2.2 :

Soient A , M , N et $A = M - N$. On dit qu'on a un splitting convergent de la matrice A si M est non singulière et si $\rho(M^{-1}N) < 1$.

Théorème 2.8 :

Soient $A = (a_{ij})$ une H -matrice et $N(A)$ la M -matrice associée à A ; si $LU = R$ est la décomposition incomplète de A associée à un ensemble $P \subseteq P_n$, et $\tilde{L}\tilde{U} = \tilde{R}$ la décomposition incomplète de $N(A)$ associée au même ensemble P alors on a : $|(LU)^{-1}R| \leq (\tilde{L}\tilde{U})^{-1}\tilde{R}$.

Preuve :

Pour démontrer ce théorème, on reprendra le procédé donné par les formules (1.2) - (1.6) pour les matrices A et $N(A)$ qu'on notera B :

$$\left\{ \begin{array}{l} A^0 = A \\ \tilde{A}_k = A_{k-1} + R_k \\ A_k = L_k \tilde{A}_k \end{array} \right\} \quad \text{pour } k = 1, \dots, n-1$$

$$\left\{ \begin{array}{l} N(A) = B = B^0 \\ \tilde{B}_k = B_{k-1} + \tilde{R}_k \\ B_k = \tilde{L}_k \tilde{B}_k \end{array} \right\} \quad \text{pour } k = 1, \dots, n-1$$

Nous ne reviendrons pas sur le fait que $\tilde{A}_k$ et A_k sont des H -matrices et que $\tilde{B}_k$ et B_k sont des M -matrices.

On arrive alors après $(n-1)$ étapes à la formule (1.6) :

$$A = LU - R = \left[\prod_{k=1}^{n-1} L_{n-k} \right]^{-1} A_{n-1} - \sum_{k=1}^{n-1} R_k$$

et

$$N(A) = B = \tilde{L} \tilde{U} - \tilde{R} = \left[\prod_{k=1}^{n-1} \tilde{L}_{n-k} \right]^{-1} B_{n-1} - \sum_{k=1}^{n-1} \tilde{R}_k$$

d'où :

$$(LU)^{-1}R = [A_{n-1}]^{-1} \left[\prod_{k=1}^{n-1} L_{n-k} \right] \left[\sum_{k=1}^{n-1} R_k \right]$$

et

$$(\tilde{L} \tilde{U})^{-1} \tilde{R} = [B_{n-1}]^{-1} \left[\prod_{k=1}^{n-1} \tilde{L}_{n-k} \right] \left[\sum_{k=1}^{n-1} \tilde{R}_k \right]$$

et on va montrer que :

$$(2.7) \quad \begin{cases} | [A_{n-1}]^{-1} | \leq [B_{n-1}]^{-1} & \text{ou} & | [A_k]^{-1} | \leq [B_k]^{-1} \\ | L_k | \leq \tilde{L}_k \\ | R_k | \leq \tilde{R}_k & \text{pour } k = 1, \dots, n-1 \end{cases}$$

en procédant étape par étape.

* Pour $k = 1$ on a :

$$A^0 = A \quad \text{et} \quad B^0 = N(A) = B$$

$$\text{donc} \quad |a_{ii}| = b_{ii} \quad \text{et} \quad |a_{ij}| = |b_{ij}| \quad \text{pour } i \neq j$$

$$\text{donc par construction :} \quad |R_1| = \tilde{R}_1$$

$$\text{et} \quad \left. \begin{array}{l} \tilde{A}_1 = A^0 + R_1 \\ \tilde{B}_1 = B^0 + \tilde{R}_1 \end{array} \right\} \quad \text{donnent} \quad \begin{cases} |\tilde{a}_{ii}| = \tilde{b}_{ii} \\ |\tilde{a}_{ij}| = -\tilde{b}_{ij} & \text{pour } i \neq j \end{cases}$$

$$\text{et par construction de } L_1 \text{ et } \tilde{L}_1 \text{ on a :} \quad |L_1| = \tilde{L}_1.$$

Les matrices $\tilde{A}_1$ et $\tilde{B}_1$ vérifient les hypothèses du théorème 2.3 donc

$$A_1 = L_1 \tilde{A}_1 \quad \text{et} \quad B_1 = \tilde{L}_1 \tilde{B}_1 \text{ vérifient les formules (2.1) :}$$

$$|a_{ii}^1| \geq b_{ii}^1 \quad \text{et} \quad -|a_{ij}^1| \geq b_{ij}^1 \text{ pour } i \neq j \text{ et on a } |(A_1)^{-1}| \leq (B_1)^{-1}.$$

* Pour $k = 2$ on a :

R_2 et $\tilde{R}_2$ par construction et du fait que $|a_{ij}^1| \leq -b_{ij}^1$ pour $i \neq j$ vérifient : $|R_2| \leq \tilde{R}_2$

$\tilde{A}_2 = A_1 + R_2$ et $\tilde{B}_2 = B_1 + \tilde{R}_2$ vérifient :

$$|\tilde{a}_{ii}^2| = |a_{ii}^1| \geq b_{ii}^1 = \tilde{b}_{ii}^2 \quad \text{et} \quad |\tilde{a}_{ij}^2| \leq -\tilde{b}_{ij}^2 \quad \text{pour } i \neq j$$

d'où :
$$\frac{|\tilde{a}_{ij}^2|}{|\tilde{a}_{ii}^2|} \leq -\frac{\tilde{b}_{ij}^2}{\tilde{b}_{ii}^2} \quad \text{donc} \quad |L_2| \leq \tilde{L}_2$$

or $\tilde{A}_2$ et $\tilde{B}_2$ vérifient les hypothèses du théorème 2.3 donc

$$A_2 = L_2 \tilde{A}_2 \quad \text{et} \quad B_2 = \tilde{L}_2 \tilde{B}_2 \quad \text{vérifient les formules (2.1) :}$$

$$|a_{ii}^2| \geq b_{ii}^2 \quad \text{et} \quad |a_{ij}^2| \leq -b_{ij}^2 \quad \text{pour } i \neq j \quad \text{qui sont encore}$$

les hypothèses du théorème 2.4, et on a $|(A_2)^{-1}| \leq (B_2)^{-1}$

et ainsi de suite pour $k = 3, \dots, n-1$ on trouve :

$$|R_k| \leq \tilde{R}_k$$

$$\tilde{A}_k = A_{k-1} + R_k \quad \text{et} \quad \tilde{B}_k = B_{k-1} + \tilde{R}_k \quad \text{qui vérifient les}$$

formules (2.1)

$$|L_k| \leq \tilde{L}_k$$

et $A_k = L_k \tilde{A}_k$, $B_k = \tilde{L}_k \tilde{B}_k$ qui vérifient les hypothèses du

théorème 2.3 et par conséquent $|(A_k)^{-1}| \leq (\tilde{B}_k)^{-1}$.

On a donc démontré les formules (2.5)

$$\text{Or} \quad |(LU)^{-1}R| \leq |U^{-1}| |L^{-1}| |R| \leq |(A_{n-1})^{-1}| \prod_{k=1}^{n-1} |L_{n-k}| \sum_{k=1}^{n-1} |R_k|$$

et compte tenu des formules (2.7), on a donc :

$$|(LU)^{-1}R| \leq (\tilde{L}\tilde{U})^{-1}\tilde{R}.$$

Or $N(A) = \tilde{L}\tilde{U} - \tilde{R}$ est un splitting régulier de la M-matrice $N(A)$ comme conséquence du théorème 1.7 donc $\rho[(\tilde{L}\tilde{U})^{-1}\tilde{R}] < 1$ comme conséquence de la proposition 1.2 ; d'où le théorème de la convergence :

Théorème 2.9 :

Soit A une H -matrice, alors pour tout $P \subseteq P_n$, la décomposition incomplète de A associée à P est un splitting convergent et la méthode itérative (1.10) converge vers l'unique solution de (1.7) pour tout x_0 .

Remarque 2.3 :

Si A est une H -matrice, $N(A)$ la M -matrice associée et si :

$A = L_1 U_1 - R_1 = L_2 U_2 - R_2$ les deux décompositions incomplètes de A associées respectivement à P_1 et P_2 , on notera la même chose pour la matrice $N(A)$:

$$N(A) = \tilde{L}_1 \tilde{U}_1 - \tilde{R}_1 = \tilde{L}_2 \tilde{U}_2 - \tilde{R}_2$$

et si $P_2 \subseteq P_1$ telles les deux décompositions de $N(A)$ soient comparables ($\tilde{R}_1 \geq \tilde{R}_2$), alors on aura :

$$\rho[(L_2 U_2)^{-1} R_2] \leq \rho[|(L_2 U_2)^{-1} R_2|] \leq \rho[|(L_1 U_1)^{-1} R_1|] \leq \rho[(\tilde{L}_1 \tilde{U}_1)^{-1} \tilde{R}_1]$$

$$\rho[(L_2 U_2)^{-1} R_2] \leq \rho[|(L_2 U_2)^{-1} R_2|] \leq \rho[(\tilde{L}_2 \tilde{U}_2)^{-1} \tilde{R}_2] \leq \rho[(\tilde{L}_1 \tilde{U}_1)^{-1} \tilde{R}_1]$$

mais en général on n'aura pas :

$$\rho[(L_2 U_2)^{-1} R_2] \leq \rho[(L_1 U_1)^{-1} R_1]$$

comme c'est le cas pour les M -matrices.

On va illustrer ceci par l'exemple suivant :

$$A = \begin{bmatrix} 3 & 1 & -2 \\ 1 & 1 & 0 \\ 0 & -1 & 2 \end{bmatrix} \quad N(A) = \begin{bmatrix} 3 & -1 & -2 \\ -1 & 1 & 0 \\ 0 & -1 & 2 \end{bmatrix}$$

et on choisit :

$$P_1 = \{(1,2), (1,3)\} \quad \text{et} \quad P_2 = \{(1,3)\}$$

et on trouve :

$$\begin{array}{lll} \rho[(L_1 U_1)^{-1} R_1] = 0 & \rho[|(L_1 U_1)^{-1} R_1|] = \frac{2}{3} & \rho[(\tilde{L}_1 \tilde{U}_1)^{-1} \tilde{R}_1] = \frac{2}{3} \\ \rho[(L_2 U_2)^{-1} R_2] = \frac{1}{2} & \rho[|(L_2 U_2)^{-1} R_2|] = \frac{1}{2} & \rho[(\tilde{L}_2 \tilde{U}_2)^{-1} \tilde{R}_2] = \frac{1}{2} \end{array}$$

Pour la stabilité numérique de la méthode des décompositions incomplètes des H-matrices, on va énoncer les théorèmes suivants :

Théorème 2.10 :

La méthode des décompositions incomplètes d'une H-matrice A est au moins aussi stable que celle des décompositions incomplètes de la M-matrice associée.

Preuve :

Considérons les décompositions incomplètes de A et N(A)

$$A = L U - R \quad \text{avec } L_{ii} = 1$$

$$N(A) = \tilde{L} \tilde{U} - \tilde{R} \quad \text{avec } \tilde{L}_{ii} = 1$$

et compte tenu du théorème 2.3 on trouve que :

$$|U_{ii}| \geq \tilde{U}_{ii} \quad \text{pour } i \in I$$

d'où la preuve du théorème.

Théorème 2.11 :

La méthode des décompositions incomplètes d'une H-matrice A est au moins aussi stable que la décomposition complète (l'élimination de Gauss) sur la M-matrice N(A) associée à A sans pivot.

Preuve :

Cela provient du théorème 2.10 et du théorème 1.8.

IV - THEOREMES DE CARACTERISATION DES H-MATRICES

Il s'agit d'énoncer quelques théorèmes de caractérisation des H-matrices qui ont un lien avec les décompositions incomplètes définies par le théorème 2.6.

Notations :

- $M_{n,n}(\mathbb{C})$: l'espace des matrices carrées d'ordre n à coefficients complexes

- $A = (a_{ij}) \in M_{n,n}(\mathbb{C})$

$$\begin{aligned} * \quad \Omega(A) &= \{ B = (b_{ij}) \in M_{n,n}(\mathbb{C}) / |b_{ij}| = |a_{ij}| \} \\ &= \{ B \in M_{n,n}(\mathbb{C}) / N(B) = N(A) \} \end{aligned}$$

$\Omega(A)$ est l'ensemble équimodulaire de A

$$* \quad \Omega^c(A) = \{ B = (b_{ij}) \in M_{n,n}(\mathbb{C}) / |b_{ii}| = |a_{ii}| \text{ et } |b_{ij}| \leq |a_{ij}| \text{ pour } i \neq j \}$$

$$* \quad \Omega^d(A) = \{ B = (b_{ij}) \in M_{n,n}(\mathbb{C}) / |b_{ii}| \geq |a_{ii}| \text{ et } |b_{ij}| \leq |a_{ij}| \text{ pour } i \neq j \}.$$

Remarquons que :

$$- \Omega(A) \subset \Omega^c(A) \subset \Omega^d(A) \quad \text{pour tout } A \in M_{n,n}(\mathbb{C})$$

- si A est une H-matrice alors $\forall B \in \Omega^d(A)$, B est une H-matrice.

On dira qu'une décomposition incomplète de A définie par le théorème 2.6 $A = LU - R$ est :

- régulière si $U_{ii} \neq 0$ pour tout $i \in I$

- positive si $U_{ii} > 0$ pour tout $i \in I$

- convergente si $(LU)^{-1}$ existe et si $\rho[(LU)^{-1}R] < 1$

- "absolument" convergente si $(LU)^{-1}$ existe et si $\rho[|(LU)^{-1}R|] < 1$.

Remarquons que :

$(LU)^{-1}$ existe $\Leftrightarrow U_{ii} \neq 0 \quad \forall i \in I$ car L est une matrice triangulaire inférieure de diagonale unitée et U est une matrice triangulaire supérieure donc $(LU)^{-1}$ existe $\Leftrightarrow$ la décomposition incomplète est régulière. On inclura la décomposition complète qui correspond à $P = \Phi$

dans les décompositions incomplètes et on conviendra que son rayon spectral est nul. On notera aussi l'ensemble des H -matrices carrées d'ordre n et à coefficients complexes par H_n .

Théorème 2.12 :

Pour tout entier $n \geq 1$

si $F_n = \{A \in M_{n,n}(\mathbb{C}) / N(A) \text{ admet une décomposition incomplète positive pour tout } P \subseteq P_n\}$

et $F_n^d = \{A \in F_n / \Omega^d(A) \subseteq F_n\}$

alors : $H_n = F_n^d$.

Preuve :

Soit $A \in H_n$ donc $N(A)$ est une M -matrice et $\forall B \in \Omega^d(A)$, $N(B)$ est aussi une M -matrice et d'après le théorème 1.5 $N(A)$ et $N(B)$ admettent une décomposition incomplète positive pour tout $P \subseteq P_n$ donc $A \in F_n^d$ et $H_n \subseteq F_n^d$. Réciproquement soit $A \in F_n^d$ donc $N(A)$ admet une décomposition incomplète positive $\forall P \subseteq P_n$ en particulier pour $P = \emptyset$ et comme $N(A)$ est une Z -matrice et d'après le théorème 1.1 $N(A)$ est une M -matrice donc $A \in H_n$ d'où $H_n = F_n^d$.

Théorème 2.13 :

Pour tout entier $n \geq 1$

si $F_n = \{A \in M_{n,n}(\mathbb{C}) / A \text{ admet une décomposition incomplète convergente pour tout } P \subseteq P_n\}$

et $F_n^d = \{A \in F_n / \Omega^d(A) \subseteq F_n\}$

alors : $H_n = F_n^d$.

Preuve :

Le théorème 2.9 montre que toute H -matrice admet une décomposition incomplète convergente pour tout $P \subseteq P_n$, donc si $A \in H_n$, $A \in F_n$

et puisque $\forall B \in \Omega^d(A) \quad B \in H_n$ donc $B \in F_n$ et par conséquent $H_n \subset F_n^d$.

Réciproquement soit $A \in F_n^d$ donc $\forall B \in \Omega^d(A)$, $B \in F_n$ en particulier :

- Pour $P = I \Rightarrow B^{-1}$ existe
- Pour $B = N(A)$ et $P = P_n$ ce qui correspond au splitting de Jacobi donc puisque $B = N(A) \in F_n \Rightarrow a_{ii} \neq 0$ et $J_B = J_{N(A)} = |J_A| \geq 0$ et on a $\rho[J_{N(A)}] = \rho[|J_A|] < 1$ donc A est une H-matrice et $H_n = F_n^d$.

Théorème 2,14 :

Pour tout entier $n \geq 1$

si $F_n = \{A \in M_{n,n}(\mathbb{C}) / A \text{ admet une décomposition incomplète "absolument" convergente pour tout } P \subseteq P_n\}$

alors $H_n = F_n$.

Preuve :

Les théorèmes 2,8 et 2,9 assurent que toute H-matrice admet une décomposition incomplète "absolument" convergente pour tout $P \subseteq P_n$ donc $H_n \subset F_n$. Réciproquement soit $A \in F_n$ donc A admet une décomposition incomplète "absolument" convergente pour tout $P \subseteq P_n$ en particulier pour $P = P_n$ ce qui correspond au splitting de Jacobi donc :

$\rho[|J_A|] < 1$ et par conséquent A est une H-matrice et $H_n = F_n$.

Théorème 2,15 :

Pour tout entier $n \geq 1$

si $F_n = \{A \in M_{n,n}(\mathbb{C}) / A \text{ admet une décomposition incomplète régulière pour tout } P \subseteq P_n\}$

et $F_n^d = \{A \in F_n / \Omega^d(A) \subseteq F_n\}$

alors : $H_n = F_n^d$.

Preuve : cf. Varga [29]

Remarques :

- Dans les théorèmes 2.12 et 2.13, on peut remplacer $\Omega^d(A)$ par $\Omega^c(A)$ ou $\Omega(A)$.
- Dans le théorème 2.15, on peut remplacer $\Omega^d(A)$ par $\Omega^c(A)$.
- Dans le théorème 2.15, si on remplace $\Omega^d(A)$ par $\Omega(A)$ et si on note $F_n' = \{A \in F_n / \Omega(A) \subseteq F_n\}$ alors on aura $F_n^d \subsetneq F_n' \subsetneq F_n$.

Théorème 2.16 :

Pour tout entier $n \geq 1$

si $F_n = \{A \in M_{n,n}(\mathbb{C}) / A \text{ admet une décomposition incomplète convergente pour tout } P \in P_n\}$

alors : $H_n \subset F_n$.

Preuve : Conséquence du théorème 2.9.

Question ouverte : Montrer que $H_n = F_n$ dans le théorème 2.16.

- CHAPITRE 3 -
=====

UNE METHODE DE DECOMPOSITIONS INCOMPLETES
D'UNE MATRICE QUELCONQUE
(NON NECESSAIREMENT UNE H-MATRICE)

Introduction :

Les décompositions incomplètes définies par les formules (1.2)-(1.5) ne peuvent être appliquées à des matrices qui ne sont pas des H-matrices (il y a des contre-exemples).

Manteuffel [13], [14] a proposé une méthode pour la décomposition incomplète d'une matrice symétrique et définie positive.

Dans ce chapitre, on reprend cette idée pour appliquer les formules de décompositions incomplètes à des matrices quelconques : on transforme la matrice initiale en une H-matrice (en lui ajoutant sa diagonale principale multipliée par un nombre réel) et on applique les décompositions incomplètes étudiées dans le chapitre précédent.

D'autres méthodes ont été proposées dans ce sens : [12], [20] pour une matrice quelconque (*), [24] pour une matrice définie positive.

(*) Cette méthode consiste à remplacer un pivot nul par un nombre non nul.

- CHAPITRE 3 -

1 - L'IDEE DE LA METHODE

Nous commençons d'abord par donner un exemple où les décompositions incomplètes définies par les formules (1.2) - (1.5) ne peuvent être appliquées à une matrice non nécessairement une H-matrice :

Exemple 3.1 :

$$A = \begin{bmatrix} 2 & 1 & \sqrt{3} \\ 1 & 2 & 1 \\ \sqrt{3} & 1 & 2 \end{bmatrix}$$

A est une matrice symétrique et définie positive mais non une H-matrice.

Prenons : $P = \{(1, 2), (2, 1)\}$

d'où :

$$L = \begin{bmatrix} 1 & & \\ 0 & 1 & \\ \sqrt{3}/2 & 1/2 & 1 \end{bmatrix} ; \quad U = \begin{bmatrix} 2 & 0 & \sqrt{3} \\ & 2 & 1 \\ & & 0 \end{bmatrix} \quad \text{et} \quad R = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

et on constate que $U_{33} = 0$, donc le procédé défini par les formules

(1.2)-(1.5) ne peut être appliqué à toute matrice.

L'idée de la méthode qui rendrait ce procédé utilisable est la suivante :

* si $A = (a_{ij})$ est une matrice carrée d'ordre n et $I = \{1, 2, \dots, n\}$

on écrira A sous la forme :

$$(3.1) \quad A = D - B$$

avec $D = (a_{ii})$ et $B = (-a_{ij})_{i \neq j}$

alors si $a_{ii} \neq 0$ pour tout $i \in I$, on peut écrire A sous la forme :

$$(3.2) \quad A = (\alpha + 1)D - B - \alpha D$$

avec α un nombre réel positif choisi de telle sorte que la matrice :

$$(3.3) \quad A(\alpha) = (\alpha + 1)D - B \text{ soit une H-matrice ;}$$

par exemple :

$$\alpha = \max_{1 \leq i \leq n} \sum_{j \neq i} |a_{ij}|$$

et au lieu d'appliquer le procédé (1.2) - (1.5) sur la matrice A , on l'appliquera sur la H-matrice $A(\alpha)$.

Et si par contre il existe $k \in I$ tel que $a_{kk} = 0$, alors on posera :

$$D' = D + Q$$

avec $D = (a_{ii})$ et $Q = (q_{ii})$ tels que :

$$\begin{cases} q_{ii} = 0 & \text{si } a_{ii} \neq 0 \\ \text{et } q_{ii} \neq 0 & \text{si } a_{ii} = 0 \end{cases}$$

alors A s'écrit sous la forme :

$$(3.4) \quad \begin{cases} A = D - B \\ = (\alpha + 1)D' - B - \alpha D' - Q \end{cases}$$

et on posera :

$$(3.5) \quad A'(\alpha) = (\alpha + 1)D' - B$$

avec α un nombre réel positif choisi tel que la matrice $A'(\alpha)$ soit une H-matrice.

Et on fera les décompositions incomplètes sur la H-matrice $A'(\alpha)$.

Remarques 3.1 :

1 - Les théorèmes étudiés dans le chapitre 2 assurent l'existence et l'unicité des décompositions incomplètes pour la H-matrice $A(\alpha)$ qu'elle soit sous la forme (3.2) ou sous la forme (3.5), pour tout $P \in P_n$.

2 - Si, pour $P \in P_n$, on écrit :

$$A(\alpha) = L(\alpha)U(\alpha) - R(\alpha)$$

$$A'(\alpha) = L'(\alpha)U'(\alpha) - R'(\alpha)$$

alors la matrice A s'écrit sous les formes :

* Pour (3.2) :

$$A = A(\alpha) - \alpha D$$

$$= L(\alpha)U(\alpha) - R''(\alpha)$$

avec $R''(\alpha) = R(\alpha) + \alpha D$

* Pour (3.4) :

$$A = A'(\alpha) - \alpha D' - Q$$

$$= L'(\alpha)U'(\alpha) - R'''(\alpha)$$

avec $R'''(\alpha) = R'(\alpha) + \alpha D' + Q.$

Les matrices $R''(\alpha)$ et $R'''(\alpha)$ n'interviennent pas puisqu'on peut les remplacer par :

$$L(\alpha)U(\alpha) - A \quad \text{ou} \quad L'(\alpha)U'(\alpha) - A.$$

3 - Quelquefois, on utilisera les formules (1.12) pour les décompositions de $A(\alpha)$:

$$(3.6) \quad A(\alpha) = L(\alpha)\Sigma(\alpha)U(\alpha) - R(\alpha).$$

4 - Si A est une H-matrice, cette méthode reste toujours valable et on constatera que α peut prendre des valeurs négatives : $\alpha > -1$ telle que $A(\alpha)$ soit une H-matrice.

II - CAS DES MATRICES DEFINIES POSITIVES

Rappelons qu'une matrice A est définie positive si :

$$(x, Ax) > 0 \quad \text{pour tout } x \neq 0.$$

Théorème 3.1 :

Soient $A = D - B$ une matrice définie positive (non nécessairement une H-matrice) et α un nombre réel positif choisi tels que la matrice $A(\alpha) = (\alpha + 1)D - B$ soit une H-matrice ; alors pour tout $P \subseteq P_n$, il existe $L(\alpha) = (L_{ij})$, $U(\alpha) = (U_{ij})$ matrices triangulaires respectivement inférieure et supérieure, $\Sigma(\alpha) = (\sigma_i)$ matrice diagonale et une matrice $R(\alpha) = (r_{ij})$ telles que :

$$(3.7) \quad \left\{ \begin{array}{l} \sigma_i > 0 \\ L_{ii} = 1 \\ U_{ii} = 1 \end{array} \right\} \quad \text{pour } i \in I$$

$$\left\{ \begin{array}{ll} L_{ij} = 0 & \text{si } (i, j) \in P \\ U_{ij} = 0 & \text{si } (i, j) \in P \\ r_{ij} = 0 & \text{si } (i, j) \notin P \end{array} \right.$$

et telles que :

$$A = L(\alpha) \Sigma(\alpha) U(\alpha) - R(\alpha).$$

Preuve :

Le théorème 2.7 nous garantit l'existence et l'unicité de la décomposition incomplète de la H-matrice $A(\alpha)$ pour un ensemble P fixé :

$$A(\alpha) = L(\alpha) \Sigma(\alpha) U(\alpha) - R'(\alpha)$$

telles que les matrices $L(\alpha)$, $U(\alpha)$, $\Sigma(\alpha)$ et $R'(\alpha)$ vérifient les formules (3.7) et on prendra $R(\alpha) = \alpha D + R'(\alpha)$ d'où :

$$\begin{aligned} A &= L(\alpha) \Sigma(\alpha) U(\alpha) - R(\alpha) \\ &= A(\alpha) - \alpha D \\ &= D - B. \end{aligned}$$

Cas particulier : Cas où A est en plus symétrique.

Théorème 3.2 :

Soient $A = D - B$ une matrice symétrique et définie positive et α un nombre réel positif choisi tels que la matrice $A(\alpha) = (\alpha + 1)D - B$ soit une H-matrice ; alors pour tout $P \subseteq P_n$ tel que si $(i, j) \in P \Rightarrow (j, i) \in P$, il existe une matrice $L(\alpha) = (L_{ij})$ triangulaire inférieure, une matrice diagonale $\Sigma(\alpha) = (\sigma_i)$ et une matrice $R(\alpha) = (r_{ij})$ tels que :

$$\left\{ \begin{array}{l} \sigma_i > 0 \\ L_{ii} = 1 \end{array} \right\} \text{ pour } i \in I$$

$$\left\{ \begin{array}{ll} L_{ij} = 0 & \text{si } (i, j) \in P \\ r_{ij} = 0 & \text{si } (i, j) \notin P \end{array} \right.$$

et telle que :

$$A = L(\alpha) \Sigma(\alpha) L^T(\alpha) - R(\alpha).$$

Preuve :

Ceci est une conséquence du théorème 2.6 et du théorème 2.7.

Remarques 3.2 :

Etant donnée une matrice A :

$$A = (a_{ij}) = D - B$$

telle que : $a_{ii} \neq 0$ pour tout $i \in I$

alors : $A(\alpha) = (\alpha + 1)D - B$ est une H-matrice si et seulement si

$$\begin{aligned}\rho[|J_{A(\alpha)}|] &= \rho\left[\left|\frac{1}{\alpha+1} D^{-1} B\right|\right] \\ &= \frac{1}{|\alpha+1|} \rho[|J_A|] < 1.\end{aligned}$$

On suppose que $\alpha > -1$ d'où :

(3.8)

$$\alpha > \rho[|J_A|] - 1$$

Alors deux cas se présentent :

1 - A est une H-matrice : donc $\rho[|J_A|] < 1$

et par conséquent : $\alpha > \rho[|J_A|] - 1$ et α peut prendre des valeurs négatives.

2 - A n'est pas une H-matrice : donc $\rho[|J_A|] \geq 1$ et $\alpha > 0$.

III - ETUDE DE LA CONVERGENCE POUR LE CAS D'UNE MATRICE DEFINIE POSITIVE

Soit à résoudre le système linéaire :

$$(3.9) \quad Ax = b$$

avec A une matrice carrée d'ordre n et définie positive, x et b deux vecteurs de $\mathbb{R}^n$.

Le théorème 3.1 donne une décomposition incomplète de la matrice A :

$$A = L(\alpha) \Sigma(\alpha) U(\alpha) - R(\alpha)$$

et on considère la méthode itérative suivante :

$$(3.10) \quad \begin{aligned} L(\alpha) \Sigma(\alpha) U(\alpha) x_{i+1} &= R(\alpha) x_i + b \\ &= [L(\alpha) \Sigma(\alpha) U(\alpha) - A] x_i + b. \end{aligned}$$

Résoudre le système (3.10) est équivalent à résoudre les systèmes :

$$(3.11) \quad \left\{ \begin{array}{ll} L(\alpha) Y_i = b - A x_i & \text{système triangulaire inférieur} \\ \Sigma(\alpha) z_i = Y_i & \text{système diagonal} \\ U(\alpha) \Delta x_i = Z_i & \text{système triangulaire supérieur} \\ \Delta x_i = x_{i+1} - x_i & \text{pour } x_0 \text{ donné et } i \geq 0. \end{array} \right.$$

Théorème 3.3 :

Soit A une matrice non singulière, $A = M - N$, avec M une matrice non singulière telle que la matrice $S = M^{-T} A^{-T} A + N$ soit définie positive alors $\rho[M^{-1}N] < 1$ si et seulement si A est définie positive.

Preuve : cf. Meurant - Golub [17] ou Raviart [22].

Cas particulier :

Si en plus A est symétrique alors $A^{-T} A = I$ et $S = M^{-T} + N$ qui est aussi symétrique.

On peut énoncer le théorème suivant :

Théorème 3.4 :

Soit A une matrice définie positive (et symétrique) alors la méthode itérative (3.11) converge vers l'unique solution de (3.9) pour tout x_0 si et seulement si la matrice

$$S = \{ [L(\alpha) \Sigma(\alpha) U(\alpha)]^T A^{-T} - I \} A + L(\alpha) \Sigma(\alpha) U(\alpha)$$

$$(S = 2 L(\alpha) \Sigma(\alpha) L^T(\alpha) - A) \text{ est définie positive.}$$

Remarques 3.3 :

$I - A$ est symétrique et définie positive, et si on considère le splitting de Jacobi :

$$\begin{aligned} A &= D - B \\ &= (\alpha + 1)D - B - \alpha D \\ &= M - N \end{aligned}$$

avec : $M = (\alpha + 1)D$ et $N = \alpha D + B$

telle que : $S = M^T + N = (2\alpha + 1)D + B$ soit définie positive

alors : $M^{-1}N = I - M^{-1}A$

$$= I - \frac{1}{\alpha + 1} D^{-1}A$$

or $D^{-1}A$ a toutes ses valeurs propres réelles et positives

et $\rho(M^{-1}N) = \max \{ |1 - \frac{1}{\alpha + 1} \lambda_{\min}(D^{-1}A)|, |1 - \frac{1}{\alpha + 1} \lambda_{\max}(D^{-1}A)| \}$

d'où si α_{opt} est la valeur optimale :

$$\alpha_{\text{opt}} = \frac{\lambda_{\max}(D^{-1}A) + \lambda_{\min}(D^{-1}A)}{2} - 1$$

2 - Si on considère le splitting de Gauss-Seidel :

$$M = (\alpha + 1)D - E \quad N = \alpha D + E^*$$

alors :

$$S = M^T + N = (2\alpha + 1)D$$

et on constate que le splitting de Gauss-Seidel converge pour tout

$$\alpha > -\frac{1}{2}$$

- CHAPITRE 4 -
=====

SUR QUELQUES METHODES ITERATIVES
DE RESOLUTION D'UN SYSTEME LINEAIRE

Introduction :

Dans ce chapitre, il est question de trois méthodes itératives de résolution d'un système linéaire : la méthode d'orthogonalisation - minimisation, la méthode du gradient conjugué généralisée et la méthode du double-gradient conjugué.

Vinsome [30] a proposé un algorithme dit d'orthogonalisation - minimisation pour résoudre un système linéaire où la matrice est une M -matrice symétrique (*). L'étude essentielle, dans ce chapitre, porte sur cet algorithme. On commence d'abord par donner une forme généralisée de l'algorithme d'orthogonalisation - minimisation et on énonce ensuite un résultat sur sa convergence.

La méthode du gradient conjugué a été utilisée sous plusieurs formes et a connu un immense succès ces dernières années ([1], [2], [5], [6], [8], [10], [18], [21], [25]); on reprend cette méthode sous sa forme généralisée proposée par Daniel [7].

On rappelle ensuite la méthode du double-gradient conjugué utilisée par Joly [32]. On ne manquera pas de montrer le lien qui existe entre ces méthodes.

Le chapitre 5 est consacré aux essais numériques de ces différentes méthodes appliquées avec les décompositions incomplètes.

(*) Vinsome n'a pas étudié la convergence de cet algorithme.

- CHAPITRE 4 -

I - NOTATIONS ET DEFINITIONS

Soit à résoudre le système linéaire :

$$(4.1) \quad Ax = b$$

avec A une matrice carrée d'ordre n à coefficients réels et non singulière x et b sont deux vecteurs de $\mathbb{R}^n$.

Définition 4.1 :

On notera $(.,.)$ le produit scalaire euclidien sur $\mathbb{R}^n \times \mathbb{R}^n$:

$$(x, y) = \sum_{i=1}^n x_i y_i = x^T y$$

et $\|.\|$ la norme euclidienne associée à ce produit scalaire :

$$\|x\| = \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}}.$$

Définition 4.2 :

On dira qu'une matrice M est définie positive si :

$$\forall x \neq 0 \quad (x, Ax) > 0.$$

On a évidemment : $(x, Ax) = 0 \Leftrightarrow x = 0$.

Définition 4.3 :

Soit M une matrice carrée d'ordre n à coefficients réels ; on notera $(.,.)_M$ le produit scalaire subordonné (s'il en est un) défini par :

(*) On notera 0 le vecteur nul de $\mathbb{R}^n$.

$$(x, y)_M = (x, My)$$

et $\| \cdot \|_M$ la norme subordonnée (si elle en est une) la norme définie par :

$$\|x\|_M = (x, Mx)^{\frac{1}{2}}.$$

Proposition 4.1 :

Soit M une matrice définie positive alors $\| \cdot \|_M$ est une norme sur $\mathbb{R}^n$.

Définition 4.4 :

Soient M une matrice réelle carrée d'ordre n et $\{p_i\}_{i \geq 0}$ une suite de vecteurs de $\mathbb{R}^n$, on dira que les éléments p_i de la suite sont :

- M -orthogonaux ou conjugués par rapport à la matrice M si

$$(p_i, p_j)_M = (p_i, Mp_j) = 0 \quad \text{pour } i \neq j$$

- M -orthogonaux à "droite" ou conjugués à "droite" par rapport à la matrice M si : $(p_i, p_j)_M = (p_i, Mp_j) = 0$ pour $i > j$.

Soit H une matrice carrée d'ordre n , réelle symétrique et définie positive. On pose $N = A^T H A$, alors N est aussi symétrique et définie positive.

Résoudre le système (4.1) est équivalent à résoudre :

$$(4.2) \quad A^T H [b - Ay] = 0 \quad \text{et si } x \text{ est la solution de 4.1 alors :}$$

$$A^T H [b - Ay] = A^T H [Ax - Ay] = 0$$

$$= A^T H A [x - y] = 0$$

$$= N(x - y) = 0.$$

Définition 4.5 :

On appellera $E(y)$ fonction erreur la forme définie par :

$$(4.3) \quad \left\{ \begin{array}{l} E(y) = (r, Hr) \quad \text{avec } r = b - Ay \\ = (b - Ay, H(b - Ay)) \\ = (A(x - y), HA(x - y)) \\ = (x - y, A^T H A(x - y)) \\ = (x - y, N(x - y)) \\ = \|x - y\|_N^2. \end{array} \right.$$

Comme N est définie positive, alors :

$$E(y) \geq 0 \quad \text{pour tout } y \text{ dans } \mathbb{R}^n \text{ et}$$

$$E(y) = 0 \Leftrightarrow y = x = A^{-1}b.$$

Remarquons que du fait de la symétrie de N on a :

$$\frac{\partial E}{\partial y_i} = -2N_{i \cdot} (x - y)$$

où $N_{i \cdot}$ est la i° ligne de la matrice N d'où :

$$\begin{aligned} \text{grad}(E(y)) &= -2N(x - y) \\ &= -2A^T H(Ax - Ay) \\ &= -2A^T H(r) \quad \text{avec } r = b - Ay \end{aligned}$$

et on a la proposition :

Proposition 4.2 :

Pour le système (4.2), le résidu en y est :

$$\rho(y) = -\frac{1}{2} \text{grad}(E(y)) = A^T Hr.$$

Proposition 4.3 :

Trouver la solution du système (4.2) c'est trouver le point où la forme $E(y)$ est minimum.

On va proposer trois méthodes itératives pour résoudre le système (4.1) en minimisant la fonction erreur $E(y)$.

Dans toute la suite, les matrices H et N désigneront les matrices définies ci-haut.

II - LA METHODE D'ORTHOGONALISATION-MINIMISATION : ORTHOMIN

Il s'agit d'une méthode qui minimise la fonction erreur $E(y)$ et qui fait intervenir deux suites de vecteurs de $\mathbb{R}^n$, $\{g_i\}$ et $\{p_i\}$ conjugués par rapport à une matrice (orthogonaux par rapport à une matrice) d'où le nom de la méthode ; elle a été proposée par P.K.W. Vinsome [30] pour résoudre (4.1) :

$$(4.4) \quad \left\{ \begin{array}{l} x_0 \text{ un vecteur arbitraire de } \mathbb{R}^n, \quad r_0 = b - Ax_0, \\ p_0 = (LDU)^{-1} r_0 \quad \text{et pour } i \geq 0 \\ \alpha_i = (r_i, Ap_i) / \|Ap_i\|^2 \\ r_{i+1} = r_i - \alpha_i Ap_i = b - Ax_{i+1} \\ a_k^{i+1} = (A(LDU)^{-1} r_{i+1}, Ap_k) / \|Ap_k\|^2 \quad \text{pour } k = 0, \dots, i \\ p_{i+1} = (LDU)^{-1} r_{i+1} - \sum_{k=0}^i a_k^{i+1} p_k \\ x_{i+1} = x_i + \alpha_i p_i \end{array} \right.$$

avec LDU une décomposition approchant la matrice A , a_k^{i+1} les paramètres de l'orthogonalisation des vecteurs Ap_i et α_i le paramètre de la minimisation de $\|r_{i+1}\|^2$. Vinsome n'a pas pu démontrer la convergence de (4.4). Nous allons généraliser cette méthode et énoncer des conditions pour qu'il y ait convergence et nous verrons que la méthode du gradient conjugué est un cas particulier de la méthode ORTHOMIN.

Proposition 4.4 :

Soit $\{p_i\}_{i \geq 0}$ une suite de vecteurs de $\mathbb{R}^{n*}$ N -orthogonaux et si on considère l'algorithme suivant :

$$(4.5) \quad \left\{ \begin{array}{l} x_0 \text{ un vecteur arbitraire de } \mathbb{R}^n \text{ et pour } i \geq 0 \\ r_i = b - Ax_i \\ c_i = (A^T H r_i, p_i) / (p_i, N p_i) \\ x_{i+1} = x_i + c_i p_i \end{array} \right.$$

alors pour tout $i \geq 0$, x_i minimise la fonction $E(y)$.

Preuve :

c_i est choisi de telle sorte que $E(x_{i+1})$ soit minimum.

Proposition 4.5 :

Soient N et K deux matrices réelles carrées d'ordre n telles que N soit symétrique et définie positive et K soit définie positive ; si on considère l'algorithme :

$$(4.6) \quad \left\{ \begin{array}{l} g_0 \in \mathbb{R}^{n*}, \quad p_0 = K g_0 \quad \text{et pour } i \geq 0 \\ c_i = (g_i, p_i) / (p_i, N p_i) \\ g_{i+1} = g_i - c_i N p_i \\ p_{i+1} = K g_{i+1} - \sum_{k=0}^i a_k^{i+1} p_k \quad \text{avec pour } k = 0, 1, \dots, i \\ a_k^{i+1} = (K g_{i+1}, N p_k) / (p_k, N p_k) \\ \text{et tels que : } (p_i, N p_j) = 0 \quad \text{pour } i \neq j \end{array} \right.$$

alors on a :

$$\begin{aligned} * \quad & (g_i, p_j) = 0 \quad \text{pour } i > j \\ * \quad & (g_i, p_i) = (g_i, K g_i) \\ * \quad & (K g_i, N p_i) = (p_i, N p_i) \end{aligned}$$

et les éléments de la suite $\{g_i\}$ sont conjugués à "droite" par rapport à la matrice K : $(g_i, K g_j) = 0$ pour $i > j$.

Preuve :

* Pour $i > j$ on a :

$$\begin{aligned} g_i &= g_{i-1} - c_{i-1} N p_{i-1} \\ &= g_j - \sum_{k=j}^{i-1} c_k N p_k \end{aligned}$$

$$\begin{aligned} \text{d'où : } (g_i, p_j) &= \left(g_j - \sum_{k=j}^{i-1} c_k N p_k, p_j \right) \\ &= (g_j, p_j) - \sum_{k=j}^{i-1} c_k (N p_k, p_j). \end{aligned}$$

Or, on a, par hypothèse :

$$(p_i, N p_j) = (N p_i, p_j) = 0 \quad \text{pour } i \neq j$$

$$\text{et } c_j = (g_j, p_j) / (p_j, N p_j)$$

d'où :

$$(g_i, p_j) = (g_j, p_j) - c_j (p_j, N p_j) = 0$$

$$\begin{aligned} * \quad (g_i, p_i) &= \left(g_i, K g_i - \sum_{k=0}^{i-1} a_k^i p_k \right) \\ &= (g_i, K g_i) - \sum_{k=0}^{i-1} a_k^i (g_i, p_k) \\ &= (g_i, K g_i) \end{aligned}$$

$$\begin{aligned} * \quad (K g_i, N p_i) &= \left(p_i + \sum_{k=0}^{i-1} a_k^i p_k, N p_i \right) \\ &= (p_i, N p_i) + \sum_{k=0}^{i-1} a_k^i (p_k, N p_i) \\ &= (p_i, N p_i) \end{aligned}$$

* Montrons que les éléments de la suite $\{g_i\}$ sont conjugués à "droite" par rapport à la matrice K ; pour $i > j$:

$$(g_i, K g_j) = (g_i, p_j) + \sum_{k=0}^{j-1} a_k^j (g_i, p_k)$$

$$\text{or on a : } (g_i, p_j) = 0 \quad \text{pour } i > j$$

$$\text{d'où : } (g_i, K g_j) = 0.$$

Remarque 4.1 :

Les égalités suivantes :

$$(K g_i, N p_i) = (p_i, N p_i)$$

$$(g_i, p_i) = (g_i, K g_i)$$

nous permettent de conclure que :

$$g_i \neq 0 \Leftrightarrow p_i \neq 0$$

et
$$g_i = 0 \Leftrightarrow p_i = 0.$$

Lemme 4.1 :

La suite $\{g_i\}$ construite par l'algorithme (4.6) s'arrête à $i = n_0$ ($\leq n$) si et seulement si $g_{n_0} = 0$.

Preuve :

Remarquons d'abord que dans l'algorithme (4.6) on calcule g_i avant p_i . Supposons qu'à l'étape i , $g_i = 0$, d'après la remarque 4.1 on aura $p_i = 0$ et on ne peut plus calculer g_{i+1} car $c_i = (g_i, p_i) / (p_i, N p_i)$ donc si $g_i = 0$ l'algorithme s'arrête. Pour montrer qu'il existe $n_0 \leq n$ tel que $g_{n_0} = 0$ nous utiliserons le lemme suivant :

Lemme 4.2 :

Soient K une matrice réelle carrée d'ordre n définie positive et $\{g_i\}_{i=0, \dots, n-1}$ une suite de vecteurs de $\mathbb{R}^{n*}$ telles que : $(g_i, K g_j) = 0$ pour $i > j$; alors $g_n = 0$.

Preuve :

g_0 étant un vecteur donné de $\mathbb{R}^{n*}$

$$Q_1 = \{g \in \mathbb{R}^n / (g, K g_0) = 0\}$$

$Q_1 \subsetneq \mathbb{R}^n$ car pour $g = g_0 \in \mathbb{R}^n$ si $g_0 \in Q_1$ on doit avoir $(g_0, K g_0) = 0$

or c'est impossible car K est une matrice définie positive et

$\dim Q_1 = n-1$ et on choisit $g_1 \in Q_1$ non nul, on notera $Q_1 = W_1$

g_2 doit vérifier :

$$(g_2, K g_0) = 0 \Rightarrow g_2 \in Q_1 = W_1$$

$$(g_2, K g_1) = 0 \Rightarrow g_2 \in Q_2 = \{g \in \mathbb{R}^n / (g, K g_1) = 0\}$$

donc

$$g_2 \in Q_1 \cap Q_2 = W_2.$$

$W_2 \subsetneq W_1$ car pour $g = g_1 \in W_1$ on doit avoir :

$$(g_1, K g_1) = 0 \quad \text{or c'est impossible car } K \text{ est définie positive}$$

et $\dim W_2 = n-2$

et ainsi de suite ; g_n doit vérifier :

$$(g_n, K g_i) = 0 \Rightarrow g_n \in Q_{i+1} \quad \text{pour } i = 0, 1, \dots, n-1$$

donc

$$g_n \in \bigcap_{i=1}^n Q_i = W_n$$

or

$$W_n \subsetneq W_{n-1} \subsetneq \dots \subsetneq W_1 \quad \text{et} \quad \dim W_n = 0 \quad \text{donc } g_n = 0.$$

Remarque 4.2 :

Pourquoi choisir la matrice K définie positive ?

Cela vient du fait que si K n'est pas définie positive, on peut trouver

un vecteur g_0 tel qu'à partir d'un certain rang $i_0, p_{i_0} = 0$ sans que

g_{i_0} soit nul et l'algorithme (4.6) s'arrête là, l'exemple suivant illustre

bien ce cas.

Exemple : où K n'est pas définie positive

$$K = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$N = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

$$\text{On part de } g_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad p_0 = K g_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad N p_0 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

$$c_0 = (g_0, p_0) / (p_0, N p_0) = 0$$

$$g_1 = g_0 - c_0 N p_0 = g_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad K g_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad a_0^1 = (K g_1, N p_0) / (p_0, N p_0) = 1$$

$$p_1 = K g_1 - a_0^1 p_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

et l'algorithme (4.6) s'arrête là sans que g_1 soit le vecteur nul.

La méthode d'orthogonalisation - minimisation pour résoudre le système linéaire (4.1) est définie par l'algorithme suivant :

Théorème 4.1 :

Soit à résoudre le système linéaire (4.1).

Soient H une matrice carrée d'ordre n réelle symétrique et définie positive et K une matrice carrée d'ordre n réelle et définie positive. On

pose $N = A^T H A$; si on considère l'algorithme suivant :

$$(4.7) \quad \left\{ \begin{array}{l} x_0 \text{ un vecteur arbitraire de } \mathbb{R}^n \\ r_0 = b - A x_0, \quad g_0 = A^T H r_0, \quad p_0 = K g_0 \quad \text{et pour } i \geq 0 \\ c_i = (g_i, p_i) / (p_i, N p_i) = (g_i, K g_i) / (p_i, N p_i) \\ g_{i+1} = g_i - c_i N p_i \\ p_{i+1} = K g_{i+1} - \sum_{k=0}^i a_k^{i+1} p_k \\ a_k^{i+1} = (K g_{i+1}, N p_k) / (p_k, N p_k) \quad \text{pour } k = 0, 1, \dots, i \\ x_{i+1} = x_i + c_i p_i \end{array} \right.$$

alors la suite $\{x_i\}$ converge vers l'unique solution x de (4.1) en n étapes au plus.

Preuve :

Remarquons d'abord que :

$$(p_i, Np_j) = 0 \quad \text{pour } i \neq j.$$

En effet, les paramètres a_k^i sont calculés de telle sorte que :

$$(p_i, Np_k) = 0 \quad \text{pour } k = 0, 1, \dots, i-1$$

et comme N est symétrique, alors on a ce résultat.

Remarquons aussi que :

$$\begin{aligned} r_{i+1} &= b - Ax_{i+1} \\ &= b - A(x_i + c_i p_i) \\ &= b - Ax_i - c_i Ap_i \\ &= r_i - c_i Ap_i. \end{aligned}$$

Multiplions les deux membres par $A^T H$ d'où :

$$A^T H r_{i+1} = A^T H r_i - c_i A^T H A p_i$$

or $g_0 = A^T H r_0$ et $A^T H A = N$ donc

$$g_{i+1} = A^T H r_{i+1} = g_i - c_i N p_i.$$

La proposition 4.5 montre que les vecteurs de la suite $\{g_i\}$ sont conjugués à "droite" par rapport à la matrice K qui est définie positive et le lemme 4.1 assure la convergence de la suite $\{g_i\}$ vers le vecteur nul en au plus n étapes, donc celle de la suite $\{r_i\}$ vers le vecteur nul et par conséquent la suite $\{x_i\}$ converge vers l'unique solution x de (4.1) en n étapes au plus.

Exemple :

Soit à résoudre le système : $Ax = b$ avec

$$A = \begin{bmatrix} 0 & 1 & -1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \quad \text{d'où } x = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

On choisit H symétrique et définie positive et K définie positive

$$H = \begin{bmatrix} 5 & -8 & 3 \\ -8 & 14 & -5 \\ 3 & -5 & 2 \end{bmatrix} \quad \text{et } K = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \quad \text{d'où } N = A^T H A = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

Soit x_0 un vecteur arbitraire de $\mathbb{R}^3$, on prend $x_0 = \begin{bmatrix} -1 \\ 0 \\ -4 \end{bmatrix}$

d'où :

$$r_0 = b - Ax_0 = \begin{bmatrix} -2 \\ 1 \\ 6 \end{bmatrix} \quad g_0 = A^T H r_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad p_0 = Kg_0 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad Np_0 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

$$c_0 = (g_0, p_0) / (p_0, Np_0) = \frac{2}{3} \quad g_1 = g_0 - c_0 Np_0 = \frac{1}{3} \begin{bmatrix} -1 \\ 1 \\ 5 \end{bmatrix} \quad Kg_1 = \frac{1}{3} \begin{bmatrix} -1 \\ 1 \\ 6 \end{bmatrix}$$

$$a_0^1 = (Kg_1, Np_0) / (p_0, Np_0) = -\frac{7}{9} \quad p_1 = Kg_1 - a_0^1 p_0 = \frac{1}{9} \begin{bmatrix} 4 \\ 10 \\ 18 \end{bmatrix} \quad Np_1 = \frac{1}{9} \begin{bmatrix} -10 \\ 10 \\ 14 \end{bmatrix}$$

$$c_1 = (g_1, p_1) / (p_1, Np_1) = \frac{12}{13} \quad g_2 = g_1 - c_1 Np_1 = \frac{1}{13} \begin{bmatrix} 9 \\ -9 \\ 3 \end{bmatrix} \quad Kg_2 = \frac{1}{13} \begin{bmatrix} 9 \\ -9 \\ -6 \end{bmatrix}$$

$$a_0^2 = (Kg_2, Np_0) / (p_0, Np_0) = \frac{5}{13} \quad a_1^2 = (Kg_2, Np_1) / (p_1, Np_1) = -\frac{99}{13^2}$$

$$p_2 = Kg_2 - a_0^2 p_0 - a_1^2 p_1 = \frac{1}{13^2} \begin{vmatrix} 96 \\ -72 \\ 120 \end{vmatrix} \quad Np_2 = \frac{1}{13^2} \begin{vmatrix} 72 \\ -72 \\ 24 \end{vmatrix}$$

$$c_2 = (g_2, p_2) / (p_2, Np_2) = \frac{13}{8} \quad g_3 = g_2 - c_2 Np_2 = \begin{vmatrix} 0 \\ 0 \\ 0 \end{vmatrix}$$

et on trouve pour la suite x_i :

$$x_1 = x_0 + c_0 p_0 = \frac{1}{3} \begin{vmatrix} -1 \\ 2 \\ -12 \end{vmatrix} \quad x_2 = x_1 + c_1 p_1 = \frac{1}{13} \begin{vmatrix} 1 \\ 22 \\ -28 \end{vmatrix}$$

$$x_3 = x_2 + c_2 p_2 = \begin{vmatrix} 1 \\ 1 \\ -1 \end{vmatrix} \quad \text{d'où } x_3 = x_*$$

Conséquences du théorème 4.1 :

* Les vecteurs x_i minimisent la fonction $E(y)$ définie par (4.3). Ceci est dû à la proposition 4.4.

* Si $g_i \neq 0$ alors on a :

$$E(x_{i+1}) = E(x_i) - c_i (g_i, Kg_i)$$

en effet :

$$\begin{aligned} E(x_{i+1}) &= (r_{i+1}, Hr_{i+1}) \\ &= (r_i - c_i Ap_i, H(r_i - c_i Ap_i)) \\ &= (r_i, Hr_i) - 2 c_i (A^T Hr_i, p_i) + c_i^2 (p_i, A^T H A p_i) \end{aligned}$$

or on a : $A^T H A = N$, $g_i = A^T H r_i$ et $c_i = (g_i, K g_i) / (p_i, N p_i)$

$$\begin{aligned} \text{d'où : } E(x_{i+1}) &= E(x_i) - c_i [2(g_i, K g_i) - c_i (p_i, N p_i)] \\ &= E(x_i) - c_i [2(g_i, K g_i) - (g_i, K g_i)(p_i, N p_i) / (p_i, N p_i)] \\ &= E(x_i) - c_i (g_i, K g_i). \end{aligned}$$

* Si on pose $\gamma(g_i) = (g_i, K g_i) / (g_i, N^{-1} g_i)$

on aura :

$$E(x_{i+1}) = E(x_0) \prod_{k=0}^i [1 - c_k \gamma(g_k)]$$

en effet, on a :

$$\begin{aligned} (g_i, N^{-1} g_i) &= (A^T H r_i, (A^T H A)^{-1} A^T H r_i) \\ &= (r_i, H A (A^T H A)^{-1} A^T H r_i) \quad H \text{ est symétrique} \\ &= (r_i, H r_i) \\ &= E(x_i) \end{aligned}$$

$$\begin{aligned} \text{d'où : } (g_i, K g_i) &= \gamma(g_i) (g_i, N^{-1} g_i) \\ &= \gamma(g_i) E(x_i) \end{aligned}$$

$$\begin{aligned} \text{et } E(x_{i+1}) &= E(x_i) - c_i (g_i, K g_i) \\ &= E(x_i) - c_i \gamma(g_i) E(x_i) \\ &= E(x_i) [1 - c_i \gamma(g_i)] \end{aligned}$$

$$\text{d'où } E(x_{i+1}) = E(x_0) \prod_{k=0}^i [1 - c_k \gamma(g_k)]$$

* Si $g_k \neq 0$ on aura :

$$0 < c_k \gamma(g_k) \leq 1$$

en effet, on a :

$$c_k = (g_k, K g_k) / (p_k, N p_k) \quad \text{et} \quad \gamma(g_k) = (g_k, K g_k) / (g_k, N^{-1} g_k)$$

et $c_k > 0$, $\gamma(g_k) > 0$ car K et N sont définies positives

d'où $c_k \gamma(g_k) > 0$

d'autre part, on a :

$$E(x_k) \geq 0$$

et $E(x_{k+1}) = E(x_k) [1 - c_k \gamma(g_k)] \geq 0$

ce qui implique que :

$$1 - c_k \gamma(g_k) \geq 0$$

donc : $0 < c_k \gamma(g_k) \leq 1$

d'où si $g_i \neq 0$:

$$\begin{aligned} E(x_{i+1}) &= E(x_0) \prod_{k=0}^i [1 - c_k \gamma(g_k)] \\ &\leq E(x_0) \left[1 - \min_{0 \leq k \leq i} (c_k \gamma(g_k))\right]^{i+1} \end{aligned}$$

ce qui peut s'écrire aussi sous la forme :

$$(4.8) \quad \|x_{i+1} - x\|_N^2 \leq \|x_0 - x\|_N^2 \left[1 - \min_{0 \leq k \leq i} (c_k \gamma(g_k))\right]^{i+1}.$$

D'où le théorème :

Théorème 4.2 :

Les éléments de la suite $\{x_i\}$ construite par l'algorithme (4.7) vérifient :

$$E(x_{i+1}) \leq E(x_0) \left[1 - \min_{0 \leq k \leq i} (c_k \gamma(g_k))\right]^{i+1}$$

pour tout $i \geq 0$ et tout x_0 .

Cas particuliers :

1 - L'algorithme (4.4) est un cas particulier de la méthode ORTHOMIN définie par l'algorithme (4.7). En effet, on choisit :

$$K = (LDU)^{-1} A^{-T} \quad \text{et} \quad H = I \quad \text{d'où} \quad N = A^T H A = A^T A.$$

Par ce choix, on remarque que la fonction $E(y)$ est le carré de la norme euclidienne du vecteur résidu $r = b - Ay$, en effet :

$$\begin{aligned} E(y) &= \|y - x\|_N^2 \\ &= \|y - x\|_{A^T A}^2 = \|A(y - x)\|^2 \end{aligned}$$

donc : $E(y) = \|Ay - b\|^2 = \|r\|^2.$

2 - Si la matrice K est en plus symétrique, alors on retrouve la méthode du gradient conjugué proposé par Daniel [7], en effet si K est symétrique, on aura :

$$(g_i, Kg_j) = 0 \quad \text{pour } i \neq j$$

et pour $k = 0, 1, \dots, i-1$ on a :

$$\begin{aligned} (Kg_{i+1}, Np_k) &= (Kg_{i+1}, \frac{1}{c_k}(g_k - g_{k+1})) \\ &= 0 \end{aligned}$$

et pour $k = i$

$$(Kg_{i+1}, Np_i) = -\frac{1}{c_i}(Kg_{i+1}, g_{i+1})$$

$$\text{d'où : } p_{i+1} = Kg_{i+1} - \sum_{k=0}^i a_k^{i+1} p_k \quad \text{avec } a_k^{i+1} = (Kg_{i+1}, Np_k) / (p_k, Np_k)$$

$$= Kg_{i+1} - a_i^{i+1} p_i$$

$$= Kg_{i+1} + a_i p_i \quad \text{avec } a_i = (g_{i+1}, Kg_{i+1}) / (g_i, Kg_i).$$

3 - Si A est symétrique et définie positive et si on choisit

$$K = I \quad \text{et} \quad H = A^{-T} \quad \text{d'où} \quad N = A \quad \text{et} \quad g_i = A^T H r_i = r_i$$

alors on retrouve la méthode classique du gradient conjugué et on aura à minimiser la fonction :

$$\begin{aligned} E(y) &= \|y - x\|_A^2 = (y - x, A(y - x)) \\ &= (y, Ay) - 2(y, b) + (x, b) \\ &= y^T A y - 2y^T b + c \quad \text{avec } c = x^T b. \end{aligned}$$

Remarque 4.3 :

La méthode d'orthogonalisation-minimisation (ORTHOMIN) définie par l'algorithme (4.7) pourrait être combinée avec les décompositions incomplètes de la matrice A (puisqu'on choisit H et K), mais en pratique, cette méthode, telle qu'elle est, ne peut être utilisée pour de grands systèmes car, pour le calcul des vecteurs p_i , on aura i orthogonalisations (par rapport à la matrice N : $(p_i, Np_k) = 0$) à faire et i paramètres d'orthogonalisation à calculer, chose qui prendrait de la place dans la mémoire de l'ordinateur et qui demanderait un temps considérable, d'où l'idée de réduire et de fixer dès le départ le nombre d'orthogonalisation à faire et l'algorithme qu'on utilisera dans la pratique sera le suivant :

$$(4.9) \quad \left\{ \begin{array}{l} m \text{ un entier qu'on fixe : nombre d'orthogonalisation à faire} \\ \text{les matrices H, K et N étant définies comme dans l'algorithme (4.7)} \\ x_0 \text{ un vecteur arbitraire} \quad r_0 = b - Ax_0 \quad g_0 = A^T H r_0 \\ p_0 = Kg_0 \quad \text{et pour } i \geq 0 \\ c_i = (g_i, p_i) / (p_i, Np_i) = (g_i, Kg_i) / (p_i, Np_i) \\ g_{i+1} = g_i - c_i Np_i = A^T H r_{i+1} \\ x_{i+1} = x_i + c_i p_i \\ p_{i+1} = Kg_{i+1} - \sum_{k=0}^{m-1} a_{i-k}^{i+1} p_{i-k} \quad \text{avec} \\ a_{i-k}^{i+1} = (Kg_{i+1}, Np_{i-k}) / (p_{i-k}, Np_{i-k}) \quad \text{pour } k = 0, 1, \dots, m-1. \end{array} \right.$$

Suivant le cas qui se présente, on utilisera des deux suites $\{r_i\}$ et $\{g_i\}$ celle qui rendrait les calculs plus faciles.

III- LA METHODE GENERALISEE DU GRADIENT CONJUGUE : CG

La méthode du gradient conjugué a été introduite par Hestenes - Stiefel [11], elle a été généralisée par Daniel [7].

Cette méthode est donnée par l'algorithme suivant :

Soit à résoudre le système (4.1)

Soient H et K deux matrices symétriques et définies positives, on pose $N = A^T H A$

$$(4.10) \quad \left\{ \begin{array}{l} x_0 \text{ un vecteur arbitraire} \quad r_0 = b - Ax_0 \quad g_0 = A^T H r_0 \\ p_0 = Kg_0 \quad \text{et pour } i \geq 0 \\ c_i = (g_i, p_i) / (p_i, N p_i) = (g_i, Kg_i) / (p_i, N p_i) \\ g_{i+1} = g_i - c_i N p_i = A^T H r_{i+1} \\ r_{i+1} = r_i - c_i A p_i \\ \alpha_i = -(Kg_{i+1}, N p_i) / (p_i, N p_i) = (g_{i+1}, Kg_{i+1}) / (g_i, Kg_i) \\ p_{i+1} = Kg_{i+1} - \alpha_i p_i \\ N_{i+1} = N_i + c_i p_i \end{array} \right.$$

On a montré que cet algorithme est un cas particulier de la méthode d'orthogonalisation - minimisation définie par l'algorithme (4.7) par le choix de la matrice K, donc l'algorithme (4.10) converge en n étapes au plus et minimise la fonction E(Y) définie par (4.3). On va énoncer des lemmes pour la majoration de cette fonction.

On pose $T = K N$ et on définit le spectre de T :

$$\text{Spect}(T) = \{\lambda \in \mathbb{C} / \det(T - \lambda I) = 0\}.$$

On remarque que si $\lambda \in \text{Spect}(T)$ alors λ est réelle positive car T est le produit de deux matrices symétriques et définies positives.

Lemme 4.3 :

On considère l'algorithme (4.10), si on pose pour $x \neq 0$:

$$u(x) = (x, N x) / (x, K^{-1} x)$$

$$v(x) = (x, K x) / (x, N^{-1} x)$$

et $a = \inf \text{Spect } (T) \quad A = \sup \text{Spect } (T)$

alors on aura :

$$(4.11) \quad \begin{cases} a \leq \mu(p_i) \leq \frac{1}{c_i} \leq \mu(Kg_i) \leq A \\ a \leq \gamma(g_i) \leq \frac{1}{c_i} \leq \gamma(Np_i) \leq A. \end{cases}$$

Preuve : cf. Daniel [7].

Proposition 4.6 :

Pour tout $i \geq 0$, on a :

$$E(x_{i+1}) \leq E(x_0) \left[1 - \frac{a}{A}\right]^{i+1}$$

et si K et N commutent, alors on a :

$$E(x_{i+1}) \leq E(x_0) \left[\frac{A-a}{A+a}\right]^{-2(i+1)}$$

Preuve :

* Compte tenu du lemme 4.3, on a pour $k = 0, 1, \dots, i$:

$$a \leq \frac{1}{c_k} \leq A \Rightarrow \frac{1}{a} \geq c_k \leq \frac{1}{A}$$

$$a \leq \gamma(g_k) \leq A$$

d'où :
$$\frac{a}{A} \leq c_k \gamma(g_k) \leq \frac{A}{a}$$

ou encore :

$$\frac{a}{A} \leq \min_{0 \leq k \leq i} [c_k \gamma(g_k)]$$

ce qui donne :

$$1 - \frac{a}{A} \geq 1 - \min_{0 \leq k \leq i} [c_k \gamma(g_k)]$$

et en appliquant la formule (4.8) :

$$E(x_{i+1}) \leq E(x_0) \left[1 - \min_{0 \leq k \leq i} [c_k \gamma(g_k)]\right]^{i+1}$$

on a alors :

$$E(x_{i+1}) \leq E(x_0) \left[1 - \frac{a}{A}\right]^{i+1}$$

qui se traduit aussi par :

$$\|x_{i+1} - x\|_N^2 \leq \|x_0 - x\|_N^2 \left[1 - \frac{a}{A}\right]^{i+1}.$$

* Si K et N commutent, alors T est symétrique et d'après le lemme 4.3 on a :

$$c_i \gamma(g_i) \geq \gamma(g_i) / \mu(Kg_i) = (g_i, Kg_i)^2 / (Kg_i, Tg_i)(Kg_i, T^{-1}g_i)$$

et on considère le produit scalaire subordonné :

$$(x, y)_{K^{-1}} = (x, K^{-1}y)$$

et on vérifie que $T = KN$ est symétrique et définie positive relativement à $(x, y)_{K^{-1}}$ avec a et A comme plus petite et plus grande valeurs propres de T, puis on utilise l'inégalité de Kantorovitch [17] et on trouve :

$$c_i \gamma(g_i) \geq \frac{4 a A}{(A + a)^2}$$

et en utilisant la formule 4.8 on a :

$$E(x_{i+1}) \leq E(x_0) \left[1 - \frac{4 a A}{(A + a)^2}\right]^{i+1} = E(x_0) \left[\frac{A - a}{A + a}\right]^{2(i+1)}$$

qui se traduit par :

$$\|x_{i+1} - x\|_N^2 \leq \|x_0 - x\|_N^2 \left[\frac{A - a}{A + a}\right]^{2(i+1)}.$$

Remarquons que la proposition 4.6 assure encore la convergence de l'algorithme (4.10).

Reprenons l'algorithme (4.10) et faisons la transformation suivante :

$$\bar{c}_i = c_i (g_i, Kg_i)$$

$$\bar{p}_i = p_i / (g_i, Kg_i)$$

on aura alors :

$$\begin{aligned}
 * \quad \bar{c}_i &= c_i(g_i, Kg_i) \\
 &= (g_i, Kg_i)^2 / (p_i, Np_i) \\
 &= 1 / (\bar{p}_i, N\bar{p}_i)
 \end{aligned}$$

$$\begin{aligned}
 * \quad g_{i+1} &= g_i - c_i Np_i \\
 &= g_i - c_i(g_i, Kg_i) N[p_i / (g_i, Kg_i)] \\
 &= g_i - \bar{c}_i N\bar{p}_i
 \end{aligned}$$

$$* \quad x_{i+1} = x_i + \bar{c}_i \bar{p}_i$$

$$\begin{aligned}
 * \quad p_{i+1} &= Kg_{i+1} + a_i p_i \quad \text{avec } a_i = (g_{i+1}, Kg_{i+1}) / (g_i, Kg_i) \\
 p_{i+1} / (g_{i+1}, Kg_{i+1}) &= Kg_{i+1} / (g_{i+1}, Kg_{i+1}) + a_i p_i / (g_{i+1}, Kg_{i+1})
 \end{aligned}$$

d'où :

$$\bar{p}_{i+1} = Kg_{i+1} / (g_{i+1}, Kg_{i+1}) + \bar{p}_i$$

et :

$$\bar{p}_0 = Kg_0 / (g_0, Kg_0)$$

et si on pose :

$$\bar{a}_{i+1} = 1 / (g_{i+1}, Kg_{i+1})$$

alors l'algorithme (4.10) s'écrit sous la forme :

$$(4.11) \quad \left\{ \begin{array}{l}
 x_0 \text{ un vecteur arbitraire} \quad r_0 = b - Ax_0 \quad g_0 = A^T H r_0 \\
 \bar{p}_0 = Kg_0 / (g_0, Kg_0) \quad \text{et pour } i \geq 0 \\
 \bar{c}_i = 1 / (\bar{p}_i, N\bar{p}_i) \\
 g_{i+1} = g_i - \bar{c}_i N\bar{p}_i = A^T H r_{i+1} \\
 r_{i+1} = r_i - \bar{c}_i A\bar{p}_i \\
 x_{i+1} = x_i + \bar{c}_i \bar{p}_i \\
 \bar{a}_{i+1} = 1 / (g_{i+1}, Kg_{i+1}) \\
 \bar{p}_{i+1} = \bar{a}_{i+1} Kg_{i+1} + \bar{p}_i.
 \end{array} \right.$$

IV - LA METHODE DU DOUBLE-GRADIENT CONJUGUE

Cette méthode est conçue surtout pour la résolution d'un système linéaire $Ax = b$ où A n'est pas une matrice symétrique définie positive mais où sa partie symétrique $(A + A^T)$ l'est. Soit A une matrice carrée d'ordre n , non singulière. On définit le système suivant :

$$B \begin{bmatrix} Y \\ x \end{bmatrix} = \begin{bmatrix} b \\ b \end{bmatrix} \quad \text{avec } B = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$$

et x , y et b des vecteurs de $\mathbb{R}^n$.

B est une matrice symétrique et non singulière.

Soit E la fonctionnelle quadratique définie par :

$$E(R) = ((R, B^{-1}R)) = (\bar{r}, A^{-1}r) + (r, A^{-T}\bar{r})$$

où $((.,.))$ désigne le produit scalaire usuel de $\mathbb{R}^{2n}$
et

$$R = \begin{bmatrix} r \\ \bar{r} \end{bmatrix} = \begin{bmatrix} b - Ax \\ b - A^T x \end{bmatrix}.$$

En développant l'expression de E en fonction de x on obtient :

$$E(R) = (x, (A + A^T)x) + \text{termes linéaires}.$$

Si la partie symétrique de A est définie positive, la fonctionnelle E est convexe, on peut donc lui appliquer un algorithme de minimisation.

L'algorithme s'écrit :

$$(4.12) \quad \left\{ \begin{array}{l} x_0 \text{ un vecteur quelconque, } r_0 = \bar{r}_0 = p_0 = \bar{p}_0 = b - Ax_0 \\ \text{pour } i \geq 0 \\ \alpha_i = (\bar{r}_i, r_i) / (\bar{p}_i, A p_i) \\ x_{i+1} = x_i + \alpha_i p_i \\ r_{i+1} = r_i - \alpha_i A p_i \\ \bar{r}_{i+1} = \bar{r}_i - \alpha_i A^T \bar{p}_i \end{array} \right.$$

$$\left\{ \begin{array}{l} \beta_i = (\bar{r}_{i+1}, r_{i+1}) / (\bar{r}_i, r_i) \\ p_{i+1} = r_{i+1} + \beta_i p_i \\ \bar{p}_{i+1} = \bar{r}_{i+1} + \beta_i \bar{p}_i \end{array} \right.$$

L'algorithme converge vers la solution du système $Ax = b$, c'est-à-dire que l'on obtient $r = 0$, mais en général $\bar{r} \neq 0$ (à l'exception du cas où A est symétrique, l'algorithme (4.12) est identique à l'algorithme classique du gradient conjugué).

On peut montrer la proposition suivante :

Proposition 4.7 :

Les suites construites par l'algorithme (4.12) vérifient :

- 1 - $(\bar{p}_i, A p_j) = (p_i, A^T \bar{p}_j) = 0 \quad \forall i > j$
- 2 - $(\bar{r}_i, r_j) = (r_i, \bar{r}_j) = 0 \quad \forall i > j$
- 3 - $\alpha_i = (\bar{r}_i, p_i) / (\bar{p}_i, A p_i) = (r_i, \bar{p}_i) / (\bar{p}_i, A p_i)$
- 4 - $\beta_i = -(r_{i+1}, A p_i) / (\bar{p}_i, A p_i) = -(r_{i+1}, A \bar{p}_i) / (\bar{p}_i, A p_i)$.

Si on fait la transformation suivante :

$$\left\{ \begin{array}{l} \alpha_i = \alpha_i (\bar{r}_i, r_i) \\ p_i^1 = p_i / (\bar{r}_i, r_i) \\ \bar{p}_i^1 = \bar{p}_i / (\bar{r}_i, r_i) \end{array} \right.$$

L'algorithme (4.12) donne :

$$\left\{ \begin{array}{l} x_0 \text{ vecteur quelconque} \quad r_0 = \bar{r}_0 = b - A x_0 \\ p_0^1 = \bar{p}_0^1 = r_0 / (\bar{r}_0, r_0) \\ \text{et pour } i \geq 0 \end{array} \right.$$

(4.13)

$$\left\{ \begin{array}{l} \alpha_i^! = 1 / (\bar{p}_i^!, A p_i^!) \\ x_{i+1} = x_i + \alpha_i^! p_i^! \\ r_{i+1} = r_i - \alpha_i^! A p_i^! \\ \bar{r}_{i+1} = \bar{r}_i - \alpha_i^! A^T \bar{p}_i^! \\ \beta_i^! = 1 / (\bar{r}_{i+1}, r_{i+1}) \\ p_{i+1}^! = \beta_i^! r_{i+1} + p_i^! \\ \bar{p}_{i+1}^! = \beta_i^! \bar{r}_{i+1} + \bar{p}_i^! \end{array} \right.$$

- CHAPITRE 5 -
=====

ESSAIS NUMERIQUES

Introduction :

Ces essais numériques portent sur trois exemples différents :

1er exemple : pour une M-matrice symétrique

2ème exemple : pour une M-matrice non symétrique

3ème exemple : pour une matrice anti-symétrique et définie positive.

Pour les deux premiers, le nombre α introduit dans le chapitre 3 n'a aucun rôle, puisqu'il s'agit de M-matrices (elles admettent des décompositions incomplètes) mais on constate qu'expérimentalement ce nombre modifie le nombre d'itérations (pour des valeurs de α négatives, le nombre d'itérations diminue). Pour le troisième exemple, la matrice n'est pas une H-matrice donc, en général, les formules des décompositions incomplètes ne sont pas applicables (on risque de rencontrer un pivot nul). Expérimentalement, ces formules restent applicables pour une telle matrice.

Les comparaisons des différentes méthodes se font par le nombre d'itérations (*), pour les différentes méthodes (gradient conjugué, double-gradient conjugué et ORTHOMIN) combinées avec différentes décompositions incomplètes et pour différents choix des matrices H et K.

On remarque que pour une matrice non symétrique (les deux derniers exemples) le nombre d'orthogonalisations pour la méthode ORTHOMIN ne dépasse pas 2 sinon on aura une précision moins bonne, alors que pour une matrice symétrique (le cas du premier exemple) le nombre d'itérations diminue si on augmente le nombre d'orthogonalisations.

Pour un problème symétrique, la méthode ORTHOMIN et la méthode du gradient conjugué combinées avec les décompositions incomplètes convergent pratiquement pour le même nombre d'itérations, mais la première demande plus de calcul et plus de place en mémoire que la dernière.

Pour un problème non symétrique, la méthode ORTHOMIN converge plus rapidement que la méthode du gradient conjugué (appliquées avec les décompositions incomplètes) et converge en autant d'itérations que la méthode du double-gradient conjugué préconditionnée, mais demande moins de calcul et moins de place en mémoire que cette dernière.

Ces essais numériques ont été réalisés sur un ordinateur MINI 6 au Centre de Calcul de Besançon et en simple précision.

(*) On avait l'intention de calculer le temps d'exécution de chaque méthode mais jusqu'à présent il n'existe pas de procédure permettant de faire cela sur l'ordinateur MINI 6.

- CHAPITRE 5 -

I - PREMIER EXEMPLE

La matrice A est une M -matrice symétrique. On considère le problème de type elliptique avec conditions de Dirichlet suivant :

$$(P) \quad \begin{cases} -\Delta u = f & \text{sur } \overset{\circ}{\Omega} \\ u = 0 & \text{sur } \Gamma = \partial\Omega \end{cases}$$

avec $\Omega = [0, 1] \times [0, 1]$.

$f(x, y) = 2[x(1-x) + y(1-y)]$ est choisie de telle sorte que

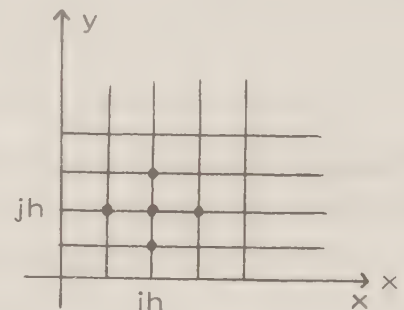
$U(x, y) = xy(1-x)(1-y)$ soit la solution du problème (P) . On définit la

grille habituelle pour $h = \frac{1}{m+1}$

$$\Omega_h = \{(ih, jh) ; 0 \leq i, j \leq m+1\}$$

$$\overset{\circ}{\Omega}_h = \{(ih, jh) ; 1 \leq i, j \leq m\}$$

$$\Gamma_h = \Omega_h \setminus \overset{\circ}{\Omega}_h$$



Soit u_h la restriction de u à Ω_h ; R_h est l'opérateur aux différences finies qui, à u_h , fait correspondre l'application définie sur $\overset{\circ}{\Omega}_h$ par :

$$\begin{aligned} R_h u_h(ih, jh) &= R_h u(i, j) \\ &= \frac{4u(i, j) - u(i-1, j) - u(i+1, j) - u(i, j-1) - u(i, j+1)}{h^2} \end{aligned}$$

en posant : $u(ih, jh) = u(i, j)$.

Au système initial (P) on associe le système :

$$(P_h) \quad \begin{cases} R_h u_h = f_h & \text{sur } \overset{\circ}{\Omega}_h \\ u_h = 0 & \text{sur } \Gamma_h. \end{cases}$$

Résoudre le système (P_h) est équivalent à résoudre un système linéaire :

$$Ax = b$$

avec : * A M-matrice symétrique carrée $m^2 \times m^2$

* $x = u_h(i, j)$ vecteur de m^2 composantes

* $b = h^2 f_h$ vecteur de m^2 composantes

$$A = \begin{array}{c} \begin{array}{c} \overbrace{\hspace{1cm}}^m \\ \left\{ \begin{array}{cccc} B & -I & & \\ -I & B & -I & \\ & -I & B & -I \\ & & -I & B \end{array} \right\} \end{array} \end{array}$$

avec :

$$B = \begin{array}{|c|c|c|c|} \hline 4 & -1 & & \\ -1 & 4 & -1 & \\ & -1 & 4 & -1 \\ & & & -1 & 4 \\ \hline \end{array}$$

I est la matrice identité.

La partie inférieure de la matrice A présente alors deux hors-diagonales non nulles en plus de la diagonale principale :

$$i - j = 1 \quad \text{et} \quad i - j = m.$$

On écrira :

$$A = D - E - F$$

et on transformera cette matrice :

$$A = (\alpha + 1)D - E - F - \alpha D$$

avec α un réel choisi tel que $A(\alpha) = (\alpha+1)D - E - F$ soit une H-matrice ;
donc si $J_A = I - D^{-1}A$ est la matrice de Jacobi associée à A et en vertu
de (3.8) on doit choisir :

$$\alpha > \rho(J_A) - 1.$$

Puis on décomposera la matrice :

$$\begin{aligned} A(\alpha) &= (\alpha+1)D - E - F \\ &= L \Sigma L^T - R \end{aligned}$$

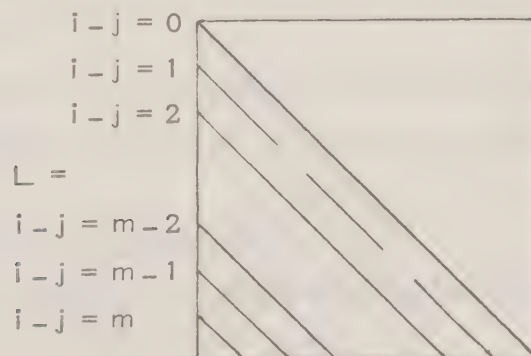
avec L matrice triangulaire inférieure et Σ matrice diagonale.

Les décompositions incomplètes qu'on propose sont des décompositions
symétriques définies par les ensembles des paires d'indices suivants :

- * $P_0 = \{(i, j) / |i-j| \neq 0, 1, m\}$
- * $P_1 = \{(i, j) / |i-j| \neq 0, 1, m-1, m\}$
- * $P_3 = \{(i, j) / |i-j| \neq 0, 1, 2, m-2, m-1, m\}$.

Remarques :

- * Les décompositions incomplètes symétriques d'une M-matrice symétrique
sont équivalentes aux décompositions incomplètes de Cholesky.
- * Les décompositions incomplètes conservent la structure bande.
- * Pour la décomposition définie par P_0 , la matrice L aura le même nom-
bre de hors-diagonales non nulles que la partie inférieure de la matrice
 $A(\alpha)$. Cette décomposition sera notée IC(0) : incomplète-Cholesky.
- * Pour P_1 la matrice L aura une hors-diagonale de plus ; elle sera notée
IC(1).
- * Pour P_3 la matrice L aura trois hors-diagonales non nulles de plus,
on la notera IC(3) :



Structure de L pour la décomposition IC(3).

La matrice A est symétrique dont seules les diagonales : $(a_{i,i})$, $(a_{i+1,i})$, $(a_{i+m,i})$, $(a_{i,i+1})$ et $(a_{i,i+m})$ sont non nulles ; donc seules les diagonales $(a_{i,i})$, $(a_{i+1,i})$ et $(a_{i+m,i})$ vont être stockées sous formes de tableaux à une dimension.

On notera : $n = m \times m$

$$(5.1) \quad \begin{cases} a_{i,i} = a_i & \text{pour } i = 1, \dots, n \\ a_{i+1,i} = \tilde{b}_i & \text{pour } i = 1, \dots, n-1 \\ a_{i+m,i} = \tilde{c}_i & \text{pour } i = 1, \dots, n-m \\ L = (L_{i,j})_{i \geq j} \\ \Sigma = (\sigma_i). \end{cases}$$

Le calcul des matrices L et Σ est donné par les formules (1.13) pour une matrice symétrique.

On peut transformer ces formules en posant : $L_{i,i} = \frac{1}{\sigma_i}$.

On aura alors :

$$(5.2) \quad \begin{cases} L_{i,i} = \frac{1}{\sigma_i} = (\alpha+1) a_{i,i} - \sum_{k=1}^{i-1} L_{i,k}^2 \sigma_k & \text{pour } i = 1, \dots, n \\ L_{i,j} = \begin{cases} a_{i,j} - \sum_{k=1}^{j-1} L_{i,k} \sigma_k L_{j,k} & \text{si } (i,j) \notin P \text{ et } i > j \\ 0 & \text{sinon.} \end{cases} \end{cases}$$

1.1 - La décomposition incomplète IC(0) :

Cette décomposition est définie par l'ensemble

$$P_0 = \{(i,j) / |i-j| \neq 0, 1, m\}.$$

Donc la matrice L aura seulement trois diagonales non nulles :

$$(L_{i,i}) ; (L_{i+1,i}) \text{ et } (L_{i+m,i}).$$

En utilisant les formules (5.2) on aura :

$$(5.3) \quad \left\{ \begin{array}{l} L_{i+m,i} = a_{i+m,i} - \sum_{k=1}^{i-1} L_{i+m,k} \sigma_k L_{i,k} \\ L_{i+1,i} = a_{i+1,i} - \sum_{k=1}^{i-1} L_{i+1,k} \sigma_k L_{i,k} \\ L_{i,i} = \frac{1}{\sigma_i} = (\alpha+1) a_{i,i} - \sum_{k=1}^{i-1} L_{i,k}^2 \sigma_k. \end{array} \right.$$

Or on vérifie que pour tout i :

$$\begin{aligned} L_{i+m,k} L_{i,k} &= 0 \\ L_{i+1,k} L_{i,k} &= 0 \end{aligned} \quad \text{pour } k = 1, \dots, i-1$$

et

$$\sum_{k=1}^{i-1} L_{i,k}^2 \sigma_k = L_{i,i-1}^2 \sigma_{i-1} + L_{i,i-m}^2 \sigma_{i-m}.$$

Les formules (5.3) donnent un calcul simple de L :

$$\left\{ \begin{array}{l} L_{i+m,i} = a_{i+m,i} \\ L_{i+1,i} = a_{i+1,i} \\ L_{i,i} = \frac{1}{\sigma_i} = (\alpha+1) a_{i,i} - L_{i,i-1}^2 \sigma_{i-1} - L_{i,i-m}^2 \sigma_{i-m} \end{array} \right.$$

et si on utilise les notations (5.1) :

$$\left\{ \begin{array}{l} L_{i+m,i} = \tilde{c}_i \\ L_{i+1,i} = \tilde{b}_i \\ L_{i,i} = \frac{1}{\sigma_i} = (\alpha+1) a_i - \tilde{b}_{i-1}^2 \sigma_{i-1} - \tilde{c}_{i-m}^2 \sigma_{i-m}. \end{array} \right.$$

On remarquera que pour cette décomposition seule la matrice diagonale Σ est à calculer.

1.2 - La décomposition incomplète IC(1) :

Cette décomposition est donnée par l'ensemble des paires d'indices

$$P_1 = \{(i,j) / |i-j| \neq 0, 1, m-1, m\}.$$

La matrice L aura alors quatre diagonales non nulles : $(L_{i,i})$; $(L_{i+1,i})$; $(L_{i+m-1,i})$ et $(L_{i+m,i})$.

En reprenant les formules (5.2), on aura :

$$(5.4) \quad \left\{ \begin{array}{l} L_{i+m,i} = a_{i+m,i} - \sum_{k=1}^{i-1} L_{i+m,k} \sigma_k L_{i,k} \\ L_{i+m-1,i} = a_{i+m-1,i} - \sum_{k=1}^{i-1} L_{i+m-1,k} \sigma_k L_{i,k} \\ L_{i+1,i} = a_{i+1,i} - \sum_{k=1}^{i-1} L_{i+1,k} \sigma_k L_{i,k} \\ L_{i,i} = \frac{1}{\sigma_i} = (\alpha+1) a_{i,i} - \sum_{k=1}^{i-1} L_{i,k}^2 \sigma_k. \end{array} \right.$$

Or l'ensemble P_1 nous impose les choses suivantes :

$$\left\{ \begin{array}{l} \sum_{k=1}^{i-1} L_{i+m,k} L_{i,k} = 0 \\ \sum_{k=1}^{i-1} L_{i+m-1,k} L_{i,k} = L_{i+m-1,i-1} L_{i,i-1} \\ \sum_{k=1}^{i-1} L_{i+1,k} L_{i,k} = L_{i+1,i-m+1} L_{i,i-m+1}. \end{array} \right.$$

Donc, en tenant compte de ceci, les formules (5.4) deviennent :

$$(5.5) \quad \left\{ \begin{array}{l} L_{i+m,i} = a_{i+m,i} \\ L_{i+m-1,i} = a_{i+m-1,i} - L_{i+m-1,i-1} \sigma_{i-1} L_{i,i-1} \\ L_{i+1,i} = a_{i+1,i} - L_{i+1,i-m+1} \sigma_{i-m+1} L_{i,i-m+1} \\ L_{i,i} = \frac{1}{\sigma_i} = (\alpha+1) a_{i,i} - L_{i,i-1}^2 \sigma_{i-1} - L_{i,i-m+1}^2 \sigma_{i-m+1} \\ \quad \quad \quad - L_{i,i-m}^2 \sigma_{i-m}. \end{array} \right.$$

Or on a $a_{i+m-1,i} = 0$, et si on pose :

$$\left\{ \begin{array}{l} L_{i+1,i} = b_i \\ L_{i+m-1,i} = h_i \end{array} \right.$$

et en utilisant les notations (5.1), les formules (5.5) deviennent :

$$\left\{ \begin{array}{l} L_{i+m,i} = \tilde{c}_i \\ L_{i+m-1,i} = h_i = -\tilde{c}_{i-1} \sigma_{i-1} b_{i-1} \\ L_{i+1,i} = b_i = \tilde{b}_i - \tilde{c}_{i-m+1} \sigma_{i-m+1} h_{i-m+1} \\ L_{i,i} = \frac{1}{\sigma_i} = (\alpha+1) a_i - b_{i-1}^2 \sigma_{i-1} - h_{i-m+1}^2 \sigma_{i-m+1} \\ \quad \quad \quad - \tilde{c}_{i-m}^2 \sigma_{i-m}. \end{array} \right.$$

On remarquera que pour cette décomposition on aura à calculer trois diagonales sous formes de vecteurs : (h_i) ; (b_i) et (σ_i) .

1.3 - La décomposition incomplète IC(3) :

Elle est définie par l'ensemble des paires d'indices :

$$P_3 = \{(i,j) / |i-j| \neq 0, 1, 2, m-2, m-1, m\}.$$

La matrice L aura donc six diagonales non nulles : $(L_{i,i})$; $(L_{i+1,i})$; $(L_{i+2,i})$; $(L_{i+m-2,i})$; $(L_{i+m-1,i})$ et $(L_{i+m,i})$.

Le calcul de ces diagonales est donné par les formules (5.2) :

$$(5.6) \quad \left\{ \begin{array}{l} L_{i+m,i} = a_{i+m,i} - \sum_{k=1}^{i-1} L_{i+m,k} \sigma_k L_{i,k} \\ L_{i+m-1,i} = - \sum_{k=1}^{i-1} L_{i+m-1,k} \sigma_k L_{i,k} \\ L_{i+m-2,i} = - \sum_{k=1}^{i-1} L_{i+m-2,k} \sigma_k L_{i,k} \\ L_{i+2,i} = - \sum_{k=1}^{i-1} L_{i+2,k} \sigma_k L_{i,k} \\ L_{i+1,i} = a_{i+1,i} - \sum_{k=1}^{i-1} L_{i+1,k} \sigma_k L_{i,k} \\ L_{i,i} = \frac{1}{\sigma_i} = (1+\alpha) a_{i,i} - \sum_{k=1}^{i-1} L_{i,k}^2 \sigma_k. \end{array} \right.$$

L'ensemble P_3 nous impose : pour $m \geq 4$

$$\left\{ \begin{array}{l} \sum_{k=1}^{i-1} L_{i+m,k} L_{i,k} = 0 \\ \sum_{k=1}^{i-1} L_{i+m-1,k} L_{i,k} = L_{i+m-1,i-1} L_{i,i-1} \\ \sum_{k=1}^{i-1} L_{i+m-2,k} L_{i,k} = L_{i+m-2,i-1} L_{i,i-1} \\ \qquad \qquad \qquad + L_{i+m-2,i-2} L_{i,i-2} \\ \sum_{k=1}^{i-1} L_{i+2,k} L_{i,k} = L_{i+2,i-m+2} L_{i,i-m+2} \\ \sum_{k=1}^{i-1} L_{i+1,k} L_{i,k} = L_{i+1,i-1} L_{i,i-1} \\ \qquad \qquad \qquad + L_{i+1,i-m+2} L_{i,i-m+2} + L_{i+1,i-m+1} L_{i,i-m+1} \end{array} \right.$$

On pose :

$$\left\{ \begin{array}{l} L_{i+1,i} = b_i \\ L_{i+2,i} = e_i \\ L_{i+m-2,i} = f_i \\ L_{i+m-1,i} = h_i \end{array} \right.$$

En utilisant ces notations avec celles du (5.1), les formules (5.6) deviennent :

$$\left\{ \begin{array}{l} L_{i+m,i} = \tilde{c}_i \\ L_{i+m-1,i} = h_i = -\tilde{c}_{i-1} \sigma_{i-1} b_{i-1} \\ L_{i+m-2,i} = f_i = -h_{i-1} \sigma_{i-1} b_{i-1} - \tilde{c}_{i-2} \sigma_{i-2} e_{i-2} \\ L_{i+2,i} = e_i = -\tilde{c}_{i-m+2} \sigma_{i-m+2} f_{i-m+2} \\ L_{i+1,i} = b_i = \tilde{b}_i - e_{i-1} \sigma_{i-1} b_{i-1} - h_{i-m+2} \sigma_{i-m+2} f_{i-m+2} \\ \qquad \qquad \qquad - \tilde{c}_{i-m+1} \sigma_{i-m+1} h_{i-m+1} \end{array} \right.$$

$$L_{i,i} = \frac{1}{\sigma_i} = (1+\alpha) a_i - b_{i-1}^2 \sigma_{i-1} - e_{i-2}^2 \sigma_{i-2} - f_{i-m+2}^2 \sigma_{i-m+2} - h_{i-m+1}^2 \sigma_{i-m+1} - \tilde{c}_{i-m} \sigma_{i-m}.$$

1.4 - Méthodes du gradient conjugué et décompositions incomplètes :

Ces méthodes seront notées ICCG(.) : Incomplete - Cholesky - Conjugate - Gradient, (.) indiquera la décomposition incomplète utilisée, soit :

- ICCG(0) pour la décomposition incomplète IC(0)
- ICCG(1) pour la décomposition incomplète IC(1)
- ICCG(3) pour la décomposition incomplète IC(3).

L'algorithme qu'on va utiliser pour ces méthodes est celui donné par les formules (4.11) :

Après avoir choisi deux matrices symétriques, définies et positives H et K, on utilisera cet algorithme :

$$(5.7) \quad \left\{ \begin{array}{l} x_0 \text{ un vecteur arbitraire } r_0 = b - Ax_0 \\ p_0 = KA^T H r_0 / (A^T H r_0, KA^T H r_0) \quad \text{et pour } i \geq 0 \\ c_i = 1 / (p_i, A^T H A p_i) \\ r_{i+1} = r_i - c_i A p_i \\ x_{i+1} = x_i + c_i p_i \\ a_{i+1} = 1 / (A^T H r_{i+1}, KA^T H r_{i+1}) \\ p_{i+1} = a_{i+1} KA^T H r_{i+1} + p_i. \end{array} \right.$$

Les programmes utilisés pour cet algorithme reposent sur l'organigramme suivant, avec calcul de nombre d'itérations (IT) :

$$x = x_0$$

$$r = b - Ax_0$$

$$p = KA^T H r / (A^T H r, KA^T H r)$$

$$IT = 0$$

$$R : IT = IT + 1$$

$$\alpha = 1 / (p, A^T H A p)$$

$$x = x + \alpha p$$

$$r = r - \alpha A p$$

IF $\|r\| \leq \epsilon$ GO TO Q

$$\beta = 1 / (A^T H r, K A^T H r)$$

$$p = \beta K A^T H r + p$$

GO TO R

Q : Print IT, x

avec : $\|r\|^2 = (r, r)$ norme euclidienne.

1.4.1 - Premier choix :

$$\left\{ \begin{array}{l} H = A^{-T} = A^{-1} \\ K = (L \Sigma L^T)^{-1}. \end{array} \right.$$

A est une M-matrice symétrique, donc A est définie positive.

Σ est une matrice diagonale à éléments strictement positifs, donc K est symétrique et définie positive.

Avec ce choix, l'algorithme (5.7) de la méthode du gradient conjugué devient :

$$\left\{ \begin{array}{l} x_0 \text{ un vecteur quelconque} \quad r_0 = b - A x_0 \\ p_0 = (L \Sigma L^T)^{-1} r_0 / (r_0, (L \Sigma L^T)^{-1} r_0) \quad \text{et pour } i \geq 0 \\ c_i = 1 / (p_i, A p_i) \\ r_{i+1} = r_i - c_i A p_i \\ x_{i+1} = x_i + c_i p_i \\ a_{i+1} = 1 / (r_{i+1}, (L \Sigma L^T)^{-1} r_{i+1}) \\ p_{i+1} = a_{i+1} (L \Sigma L^T)^{-1} r_{i+1} + p_i. \end{array} \right.$$

On remarquera que pour ce choix il y a peu de calcul à faire et qu'on a à minimiser la fonction :

$$E(x_i) = \|x_i - x\|_A^2 = (x_i - x, A(x_i - x)).$$

On remarquera aussi que si on pose $T = K A^T H A = (L \Sigma L^T)^{-1} A$, alors toutes les valeurs propres de T sont réelles et positives, et si $\lambda_{\min}(T) = \inf \text{Spect}(T)$ et $\lambda_{\max}(T) = \sup \text{Spect}(T)$, on aura une majoration de la fonction $E(x_i)$ à la $i^{\text{ème}}$ itération :

$$E(x_i) \leq E(x_0) \left[1 - \frac{\lambda_{\min}(T)}{\lambda_{\max}(T)} \right]^i$$

ou bien :

$$\|x_i - x\|_A^2 \leq \|x_0 - x\|_A^2 \left[1 - \frac{\lambda_{\min}(T)}{\lambda_{\max}(T)} \right]^i.$$

Cette majoration est donnée par la proposition 4.6.

Les essais numériques pour ce choix sont réalisés pour $m = 19$, $x_0 = 0$ (vecteur nul) avec un test d'arrêt $\epsilon = 10^{-6}$.

Les tableaux suivants donnent le nombre d'itérations (IT) et $\text{Log}_{10} \|r\|$ pour les différentes valeurs du réel α (Alpha) et les différentes décompositions incomplètes.

ICCG1(0), ICCG1(1), ICCG1(3) : seront les méthodes Incomplete - Cholesky - Conjugate - Gradient pour le premier choix des matrices H et K appliquées respectivement aux décompositions IC(0), IC(1) et IC(3).

Tableau I.1 :

Méthodes du gradient conjugué combinées avec les décompositions incomplètes IC(.) pour le choix suivant :

$$\begin{cases} H = A^{-1} \\ K = (L \Sigma L^T)^{-1}. \end{cases}$$

Elles seront notées ICCG (.)

m = 19 $\epsilon = 10^{-6}$

	ICCG 1(0)		ICCG 1(1)		ICCG 1(3)	
Alpha	IT	$\text{Log}_{10} \ r\ $	IT	$\text{Log}_{10} \ r\ $	IT	$\text{Log}_{10} \ r\ $
0.025	14	-6.165	10	-6.513	8	-6.460
0.020	14	-6.101	10	-6.512	7	-6.081
0.015	14	-6.031	10	-6.546	7	-6.268
0.010	15	-6.754	9	-6.042	7	-6.421
0.005	14	-6.275	9	-6.147	7	-6.648
0.000	14	-6.691	9	-6.264	7	-6.867
-0.005	13	-6.056	9	-6.373	7	-7.032
-0.010	13	-6.211	9	-6.487	7	-6.892
-0.015	13	-6.309	9	-6.594	7	-6.921
-0.020	13	-6.350	8	-6.046	6	-6.037
-0.025	13	-6.366	8	-6.116	6	-6.116

avec $r = b - Ax$.

1.4.2 - Le deuxième choix :

$$\left\{ \begin{array}{l} H = (A A^T)^{-1} = (A A)^{-1} \\ K = A^T (L \Sigma L^T)^{-1} (L \Sigma L^T)^{-1} A = A (L \Sigma L^T)^{-1} (L \Sigma L^T)^{-1} A. \end{array} \right.$$

Les matrices H et K sont symétriques et définies positives.

Par ce choix l'algorithme (5.7) devient :

$$\left\{ \begin{array}{l} x_0 \text{ un vecteur arbitraire} \quad r_0 = b - A x_0 \\ p_0 = A (L \Sigma L^T)^{-1} (L \Sigma L^T)^{-1} r_0 / (r_0, (L \Sigma L^T)^{-1} (L \Sigma L^T)^{-1} r_0) \\ \text{et pour } i \geq 0 \\ c_i = 1 / (p_i, p_i) = 1 / \|p_i\|^2 \\ r_{i+1} = r_i - c_i A p_i \\ x_{i+1} = x_i + c_i p_i \\ a_{i+1} = 1 / (r_{i+1}, (L \Sigma L^T)^{-1} (L \Sigma L^T)^{-1} r_{i+1}) \\ p_{i+1} = a_{i+1} A (L \Sigma L^T)^{-1} (L \Sigma L^T)^{-1} r_{i+1} + p_i. \end{array} \right.$$

Cet algorithme devient encore, si on pose $s_i = (L \Sigma L^T)^{-1} r_i$:

$$\left\{ \begin{array}{l} x_0 \text{ un vecteur arbitraire} \quad s_0 = (L \Sigma L^T)^{-1} [b - A x_0] \\ p_0 = A (L \Sigma L^T)^{-1} s_0 / \|s_0\|^2 \\ \text{et pour } i \geq 0 \\ c_i = 1 / \|p_i\|^2 \\ s_{i+1} = s_i - c_i (L \Sigma L^T)^{-1} A p_i \\ x_{i+1} = x_i + c_i p_i \\ a_{i+1} = 1 / \|s_{i+1}\|^2 \\ p_{i+1} = a_{i+1} A (L \Sigma L^T)^{-1} s_{i+1} + p_i. \end{array} \right.$$

Ce choix de H et K nous permet aussi de minimiser la fonction :

$$E(x_i) = \|x_i - x\|^2 = (x_i - x, x_i - x).$$

On remarquera que pour ce deuxième choix la programmation de l'algorithme demandera plus de calcul et plus de place en mémoire que pour le premier choix. Si on pose :

$$T = K A^T H A = A (L \Sigma L^T)^{-1} (L \Sigma L^T)^{-1} A.$$

On remarquera que T est symétrique et on aura une majoration de la fonction $E(x_i)$ donnée par la proposition 4.6 :

$$E(x_i) \leq E(x_0) \left[\frac{\lambda_{\max}(T) - \lambda_{\min}(T)}{\lambda_{\max}(T) + \lambda_{\min}(T)} \right]^{2i}$$

ou encore :

$$\|x_i - x\|^2 \leq \|x_0 - x\|^2 \left[\frac{\lambda_{\max}(T) - \lambda_{\min}(T)}{\lambda_{\max}(T) + \lambda_{\min}(T)} \right]^{2i}.$$

Les essais numériques pour ce deuxième choix sont réalisés pour $m = 19$, $x_0 = 0$ et une précision $\epsilon = 10^{-5}$.

Le tableau 3 donne le nombre d'itérations (IT) et $\text{Log}_{10} \|r\|$ pour les différentes valeurs de α (Alpha) et les différentes décompositions incomplètes.

ICCG2(0), ICCG2(1) et ICCG2(3) sont les méthodes Incomplete - Cholesky Conjugate - Gradient pour le deuxième choix des matrices H et K, appliquées respectivement aux décompositions incomplètes IC(0), IC(1) et IC(3).

Tableau 1.2 : $K = A (L \Sigma L^T)^{-1} (L \Sigma L^T)^{-1} A$ $H = (A A)^{-1}$

$m = 19$ $\epsilon = 10^{-5}$

	ICCG 2(0)		ICCG 2(1)		ICCG 2(3)	
Alpha	IT	$\log_{10} \ b - Ax\ $	IT	$\log_{10} \ b - Ax\ $	IT	$\log_{10} \ b - Ax\ $
0.0120	26	- 5.3896923	15	- 5.0251970	10	- 5.7103624
0.0096	27	- 5.6120272	15	- 5.5593662	9	- 5.1249723
0.0072	26	- 5.3976192	14	- 5.0823374	9	- 5.3023148
0.0048	27	- 5.4176493	14	- 5.2700472	9	- 5.4214315
0.0024	26	- 5.5472527	14	- 5.3458233	9	- 5.5252590
0.0000	26	- 5.1191092	14	- 5.3993797	8	- 5.0934992
- 0.0024	25	- 5.1295757	14	- 5.4648819	8	- 5.2343769
- 0.0048	25	- 5.1710367	13	- 5.0416107	8	- 5.3920498
- 0.0072	26	- 5.3294601	13	- 5.1449728	8	- 5.3128557
- 0.0096	25	- 5.3199902	13	- 5.2860460	8	- 5.3088140
- 0.0120	25	- 5.1100807	13	- 5.4016590	8	- 5.3370037

1.5 - Méthodes d'orthogonalisation - minimisation et décompositions incomplètes :

Ces méthodes seront notées ORTHOMIN $k(.)$:

ORTHOMIN : ORTHOGonalisation - MINimisation

k : le nombre d'orthogonalisations à faire

$(.)$: indiquera la décomposition incomplète utilisée.

L'algorithme utilisé est celui donné par les formules (4.9) :

$$\left\{ \begin{array}{l} k \text{ un entier que l'on fixe : nombre d'orthogonalisations à faire ;} \\ H \text{ une matrice symétrique et définie positive} \\ K \text{ une matrice définie positive (non symétrique)} \\ x_0 \text{ un vecteur arbitraire} \quad r_0 = b - Ax_0 \quad p_0 = KA^T H r_0 \\ \text{et pour } i \geq 0 \\ c_i = (A^T H r_i, KA^T H r_i) / (p_i, A^T H A p_i) \\ r_{i+1} = r_i - c_i A p_i \\ x_{i+1} = x_i + c_i p_i \\ a_{i-j}^{i+1} = (KA^T H r_{i+1}, A^T H A p_{i-j}) / (p_{i-j}, A^T H A p_{i-j}) \\ \text{pour } j = 0, 1, \dots, k-1 \\ p_{i+1} = KA^T H r_{i+1} - \sum_{j=0}^{k-1} a_{i-j}^{i+1} p_{i-j}. \end{array} \right.$$

Les programmes utilisés pour cet algorithme reposeront sur l'organigramme suivant avec calcul de nombre d'itérations (IT) :

$$\left\{ \begin{array}{l} x = x_0 \\ r = b - Ax \\ p = KA^T H r \\ p^j = 0 \quad \text{pour } j = 1, 2, \dots, k \\ IT = 0 \\ R : IT = IT + 1 \end{array} \right.$$

$$\alpha = (A^T H r, K A^T H r) / (p, A^T H A p)$$

$$x = x + \alpha p$$

$$r = r - \alpha A p$$

IF $\|r\| \leq \epsilon$ GO TO Q

CHANGE

CALCUL

$$p = K A^T H r - s$$

GO TO R

Q : Print x, IT

avec :

CHANGE : un sous-programme qui change p^j en p^{j-1} pour $j = 2, 3, \dots, k$ et p^1 en p .

CALCUL : un sous-programme qui calcule les coefficients a_{i-j}^{i+1} pour $j = 1, 2, \dots, k$ et la somme $s = \sum_{j=1}^k a_{i-j}^{i+1} p^j$.

Choix des matrices H et K :

Pour ces méthodes, nous n'utiliserons qu'un seul choix :

$$\begin{cases} H = I \\ K = (L \Sigma L^T)^{-1} A^{-T} = (L \Sigma L^T)^{-1} A^{-1}. \end{cases}$$

Par ce choix, l'algorithme (5.8) devient :

$$\left\{ \begin{array}{l} k \text{ entier fixé, } x_0 \text{ vecteur arbitraire} \\ r_0 = b - A x_0 \quad p_0 = (L \Sigma L^T)^{-1} r_0 \\ \text{et pour } i \geq 0 \\ c_i = (r_i, A (L \Sigma L^T)^{-1} r_i) / (p_i, A A p_i) \\ \quad = (r_i, A (L \Sigma L^T)^{-1} r_i) / \|A p_i\|^2 \end{array} \right.$$

$$\left\{ \begin{array}{l} r_{i+1} = r_i - c_i A p_i \\ x_{i+1} = x_i + c_i p_i \\ \alpha_{i+1}^{LS} = \frac{(A(L \Sigma L^T)^{-1} r_{i+1})^T A p_{i-1}}{\|A p_{i-1}\|^2} \\ p_{i+1} = (L \Sigma L^T)^{-1} r_{i+1} - \sum_{j=0}^{i-1} \alpha_{i-j}^{i+1} p_{i-j} \end{array} \right.$$

Cet algorithme converge si la matrice $K = (L \Sigma L^T)^{-1} A^{-1}$ est définie positive. La matrice $(L \Sigma L^T)^{-1}$ approche A^{-1} , donc on peut mettre K sous la forme :

$$\begin{aligned} K &= (L \Sigma L^T)^{-1} A^{-1} \\ &= (L \Sigma L^T)^{-1} A A^{-1} A^{-1} \\ &= A^{-1} A^{-1} \end{aligned}$$

Or A est symétrique, donc K reste proche d'une matrice symétrique et définie positive.

Remarque importante :

- Le choix de $K = (L \Sigma L^T)^{-1} A^{-1}$ pour cette méthode est dû au fait que si $K = A^{-1} A^{-1}$, l'algorithme d'orthogonalisation - minimisation converge en une itération, or $(L \Sigma L^T)^{-1}$ est une décomposition approchant A^{-1} , donc par ce choix, on espère ressembler plus près de l'algorithme qui converge en une itération.

Tableaux 1.3 :

On présente trois tableaux donnant les résultats de la méthode d'orthogonalisation - minimisation pour les trois décompositions incomplètes IC(.) et le choix suivant :

$$\begin{cases} H = I \\ K = (L \Sigma L^T)^{-1} A^{-1}. \end{cases}$$

Elle est notée ORTHOMIN k(.) avec k le nombre d'orthogonalisations et (.) représente la décomposition utilisée, k prend les valeurs 1, 2, 3, 4 et 5

$$m = 19 \quad \epsilon = 10^{-6}$$

Tableau 1.3.1 : Pour la décomposition IC(0)

Alpha	ORTHOMIN 1(0)		ORTHOMIN 2(0)		ORTHOMIN 3(0)		ORTHOMIN 4(0)		ORTHOMIN 5(0)	
	IT	$\text{Log}_{10} r $	IT	$\text{Log}_{10} r $	IT	$\text{Log}_{10} r $	IT	$\text{Log}_{10} r $	IT	$\text{Log}_{10} r $
0.025	16	-6.090	15	-6.060	15	-6.120	15	-6.161	14	-6.139
0.020	16	-6.151	15	-6.080	15	-6.141	15	-6.183	14	-6.147
0.015	15	-6.007	15	-6.099	15	-6.159	14	-6.000	14	-6.176
0.010	15	-6.027	15	-6.124	15	-6.177	14	-6.032	13	-6.001
0.005	15	-6.040	15	-6.158	15	-6.195	14	-6.055	13	-6.008
0.000	15	-6.055	15	-6.190	15	-6.214	14	-6.081	13	-6.041
-0.005	15	-6.079	15	-6.213	15	-6.236	14	-6.099	13	-6.058
-0.010	15	-6.114	15	-6.234	14	-6.012	14	-6.116	13	-6.071
-0.015	15	-6.159	15	-6.261	14	-6.031	14	-6.131	13	-6.092
-0.020	15	-6.205	15	-6.295	14	-6.056	14	-6.150	13	-6.120
-0.025	15	-6.247	14	-6.013	14	-6.091	14	-6.159	13	-6.127

Tableau 1.3.2 : Pour la décomposition IC(1)

	ORTHOMIN 1(1)		ORTHOMIN 2(1)		ORTHOMIN 3(1)		ORTHOMIN 4(1)		ORTHOMIN 5(1)	
	IT	$\text{Log}_{10} \ r\ $	IT	$\text{Log}_{10} \ r\ $	IT	$\text{Log}_{10} \ r\ $	IT	$\text{Log}_{10} \ r\ $	IT	$\text{Log}_{10} \ r\ $
Alpha										
0.025	11	-6.225	10	-6.195	9	-6.086	10	-6.249	9	-6.061
0.020	11	-6.305	10	-6.103	9	-6.129	9	-6.288	9	-6.100
0.015	10	-6.100	10	-6.069	9	-6.170	9	-6.042	9	-6.142
0.010	9	-6.246	10	-6.074	9	-6.229	9	-6.122	9	-6.218
0.005	10	-6.149	10	-6.102	9	-6.306	9	-6.238	9	-6.321
0.000	10	-6.122	10	-6.147	9	-6.374	9	-6.342	9	-6.404
-0.005	10	-6.181	10	-6.216	9	-6.443	9	-6.472	9	-6.476
-0.010	10	-6.275	10	-6.315	9	-6.471	9	-6.606	9	-6.547
-0.015	10	-6.356	10	-6.402	8	-6.031	8	-6.081	8	-6.045
-0.020	10	-6.417	10	-6.464	8	-6.037	8	-6.124	8	-6.075
-0.025	10	-6.504	9	-6.002	8	-6.042	8	-6.158	8	-6.093

Tableau 1.3.3 : Pour la décomposition IC(3)

Alpha	ORTHOMIN 1(3)		ORTHOMIN 2(3)		ORTHOMIN 3(3)		ORTHOMIN 4(3)		ORTHOMIN 5(3)	
	IT	$\text{Log}_{10} \ r\ $	IT	$\text{Log}_{10} \ r\ $	IT	$\text{Log}_{10} \ r\ $	IT	$\text{Log}_{10} \ r\ $	IT	$\text{Log}_{10} \ r\ $
0.025	9	-6.436	8	-6.488	8	-6.642	8	-6.643	8	-6.595
0.020	8	-6.003	7	-6.167	7	-6.177	7	-6.152	7	-6.094
0.015	8	-6.250	7	-6.350	7	-6.357	7	-6.295	7	-6.153
0.010	7	-6.505	7	-6.339	7	-6.491	7	-6.358	7	-6.133
0.005	7	-6.052	7	-6.217	7	-6.623	7	-6.381	7	-6.144
0.000	8	-6.714	7	-6.150	7	-6.635	7	-6.384	7	-6.176
-0.005	7	-6.024	7	-6.219	7	-6.577	7	-6.431	7	-6.191
-0.010	7	-6.062	7	-6.392	7	-6.732	7	-6.846	7	-6.740
-0.015	7	-6.249	7	-6.821	7	-6.869	6	-6.003	7	-6.800
-0.020	7	-6.432	7	-6.482	6	-6.131	6	-6.122	6	-6.055
-0.025	7	-6.583	7	-6.536	6	-6.160	6	-6.163	6	-6.100

Conclusions : pour un problème symétrique

- * La méthode du gradient conjugué utilisée pour le premier choix $(H = A^{-1} \text{ et } K = (L \Sigma L^T)^{-1})$ converge mieux que cette méthode pour un autre choix,
- * La méthode ORTHOMIN converge en un nombre d'itérations pratiquement égal à celui de la méthode du gradient conjugué pour le premier choix, mais cette dernière demande moins de calcul et moins de place en mémoire que la première,
- * Toutes les méthodes convergent mieux pour une décomposition incomplète qui s'approche de la décomposition complète.
Il est donc conseillé d'utiliser, pour des problèmes symétriques, la méthode du gradient conjugué pour le premier choix et pour une décomposition incomplète plus proche de la décomposition complète.

II - DEUXIEME EXEMPLE

La matrice A est une M-matrice non symétrique. On considère le problème de type elliptique, avec conditions de Newmann, suivant :

$$(P) \begin{cases} -\Delta u + wu = f & \text{sur } \overset{\circ}{\Omega} \\ \frac{\partial u}{\partial n/r} = 0 & r = \partial\Omega \end{cases}$$

avec $\Omega = [0, 1] \times [0, 1]$

w une fonction strictement positive

et $\frac{\partial u}{\partial n}$ la dérivée normale de u .

On utilisera le maillage de Ω défini dans le premier exemple.

L'opérateur R_h aux différences finies sera défini par :

$$R_h u_{i,j} = \frac{1}{h^2} [(4 + w_{ij} h^2) u_{i,j} - u_{i-1,j} - u_{i+1,j} - u_{i,j-1} - u_{i,j+1}].$$

Au problème (P) on associe le problème suivant :

$$(P_h) \begin{cases} R_h U_h = f_h & \text{sur } \overset{\circ}{\Omega}_h \\ u_{i-1,j} = u_{i+1,j} & \text{et } u_{i,j-1} = u_{i,j+1} & \text{sur } r_h. \end{cases}$$

Le système linéaire $Ax = b$ associé au problème (P_h) est tel que :

$b = (h^2 f_{i,j})$ et A la matrice carrée $m^2 \times m^2$ suivante :

$$A = \begin{array}{|c|c|c|c|} \hline B & -2I & & \\ \hline -I & B & -I & \\ \hline & -I & B & -I \\ \hline & & -2I & B \\ \hline \end{array}$$

avec B une matrice carrée $m \times m$ de la forme :

$$B = \begin{array}{ccccc} & & & & \\ & & & & \\ 4+h^2_{\omega} & & -2 & & \\ & -1 & & 4+h^2_{\omega} & -1 \\ & & & & \\ & & & & \\ & & & & \\ & & -1 & & 4+h^2_{\omega} & -1 \\ & & & & & \\ & & & -2 & & 4+h^2_{\omega} \end{array}$$

On prendra $\omega =$ constante positive.

Les décompositions incomplètes que l'on propose sont les décompositions définies par les ensembles des paires d'indices suivants :

$$* P_0 = \{(i, j) / |i - j| \neq 0, 1, m\}$$

$$* P_3 = \{(i, j) / |i - j| \neq 0, 1, 2, m-2, m-1, m\}.$$

Ces décompositions incomplètes seront de deux natures :

- décompositions incomplètes symétriques : on transformera la matrice A en une matrice symétrique et on utilisera les décompositions sur cette dernière
- décompositions incomplètes "à positions symétriques" : on appliquera les formules des décompositions incomplètes directement sur la matrice A.

II.1 - Les décompositions incomplètes symétriques :

On transformera la matrice A en une somme de deux matrices $\tilde{A}$ et N de telle sorte que la matrice $\tilde{A}$ soit celle trouvée dans le premier exemple et on appliquera les formules des décompositions incomplètes sur la matrice $\tilde{A}$.

Ces décompositions ont été faites au premier exemple : I.2 et I.3, elles seront notées comme au premier exemple : IC(0), IC(1) et IC(3).

II.2 - Les décompositions incomplètes "à positions symétriques" :

Ces décompositions incomplètes seront appliquées directement sur la matrice A en appliquant les formules. Elles seront notées DI(.)

si $A = D - E - F$ on écrira aussi

$$A = (\alpha+1)D - E - F - \alpha D$$

et on fera les décompositions sur la matrice $(\alpha+1)D - E - F$.

La matrice A s'écrira alors :

$$A = L \Sigma U - R - \alpha D \quad \text{si } (\alpha+1)D - E - F = L \Sigma U - R.$$

Le calcul des matrices L , Σ et U est donné par les formules (1.12) ;

on peut transformer ces formules en posant : $L_{ii} = U_{ii} = \frac{1}{\sigma_i}$. On aura :

$$(5.9) \quad \left\{ \begin{array}{l} \frac{1}{\sigma_i} = L_{i,i} = U_{i,i} = (\alpha+1)a_{i,i} - \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i} \\ \\ L_{i,j} = \begin{cases} a_{i,j} - \sum_{k=1}^{j-1} L_{i,k} \sigma_k U_{k,j} & \text{si } (i,j) \notin P \text{ et } i > j \\ 0 & \text{sinon} \end{cases} \\ \\ U_{i,j} = \begin{cases} a_{i,j} - \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,j} & \text{si } (i,j) \notin P \text{ et } j > i \\ 0 & \text{sinon} \end{cases} \end{array} \right.$$

avec : $A = (a_{i,j})$, $\Sigma = (\sigma_i)$, $L = (L_{i,j})_{i \geq j}$ et $U = (U_{i,j})_{j \geq i}$.

II.2.1 - La décomposition incomplète DI(0) :

Cette décomposition est définie par l'ensemble P_0 , les matrices

L et U auront le même nombre de hors-diagonales que les parties correspondantes de la matrice A ; les formules (5.9) nous donnent le calcul effectif de ces matrices :

$$\left\{ \begin{array}{l} L_{i+1,i} = a_{i+1,i} - \sum_{k=1}^{i-1} L_{i+1,k} \sigma_k U_{k,i} \\ \\ L_{i+m,i} = a_{i+m,i} - \sum_{k=1}^{i-1} L_{i+m,k} \sigma_k U_{k,i} \\ \\ U_{i,i+1} = a_{i,i+1} - \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i+1} \end{array} \right.$$

$$\left\{ \begin{array}{l} U_{i,i+m} = a_{i,i+m} - \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i+m} \\ L_{i,i} = U_{i,i} = \frac{1}{\sigma_i} = (\alpha+1) a_{i,i} - \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i} \end{array} \right.$$

Or l'ensemble P_0 impose que :

$$\left\{ \begin{array}{l} L_{i+1,k} U_{k,i} = 0 \\ L_{i+m,k} U_{k,i} = 0 \\ L_{i,k} U_{k,i+1} = 0 \\ L_{i,k} U_{k,i+m} = 0 \end{array} \right\} \text{ pour } k = 1, \dots, i-1$$

$$\sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i} = L_{i,i-1} \sigma_{i-1} U_{i-1,i} + L_{i,i-m} \sigma_{i-m} U_{i-m,i}$$

d'où :

$$\left\{ \begin{array}{l} L_{i+1,i} = a_{i+1,i} \\ L_{i+m,i} = a_{i+m,i} \\ U_{i,i+1} = a_{i,i+1} \\ U_{i,i+m} = a_{i,i+m} \\ L_{i,i} = U_{i,i} = \frac{1}{\sigma_i} = (\alpha+1) a_{i,i} - L_{i,i-1} \sigma_{i-1} U_{i-1,i} \\ \quad - L_{i,i-m} \sigma_{i-m} U_{i-m,i} \end{array} \right.$$

Or on introduira les éléments de la matrice A sous forme de tableaux à une dimension :

$$(5.10) \quad \left\{ \begin{array}{l} a_{i,i} = a_i \\ a_{i+1,i} = b_i \\ a_{i+m,i} = c_i \\ a_{i,i+1} = b_i' \\ a_{i,i+m} = c_i' \end{array} \right.$$

En adoptant ces nouvelles notations, les formules de calcul des deux matrices L et U deviennent :

$$\left\{ \begin{array}{l} L_{i+1,i} = b_i \\ L_{i+m,i} = c_i \\ U_{i,i+1} = b'_i \\ U_{i,i+m} = c'_i \\ L_{i,i} = U_{i,i} = \frac{1}{\sigma_i} = (\alpha+1)a_i - b_{i-1}\sigma_{i-1}b'_{i-1} - c_{i-m}\sigma_{i-m}c'_{i-m} \end{array} \right.$$

On remarquera que pour cette décomposition on ne calculera que les éléments de la matrice diagonale $\Sigma = (\sigma_i)$.

II.2.2 - La décomposition incomplète DI(3) :

Cette décomposition est définie par l'ensemble P_3 , les matrices L et U auront chacune trois hors-diagonales de plus que les parties inférieure et supérieure de la matrice A.
Le calcul effectif de ces matrices est donné par les formules (5.9) qui se traduisent par :

$$\left\{ \begin{array}{l} L_{i+1,i} = a_{i+1,i} - \sum_{k=1}^{i-1} L_{i+1,k} \sigma_k U_{k,i} \\ L_{i+2,i} = a_{i+2,i} - \sum_{k=1}^{i-1} L_{i+2,k} \sigma_k U_{k,i} \\ L_{i+m-2,i} = a_{i+m-2,i} - \sum_{k=1}^{i-1} L_{i+m-2,k} \sigma_k U_{k,i} \\ L_{i+m-1,i} = a_{i+m-1,i} - \sum_{k=1}^{i-1} L_{i+m-1,k} \sigma_k U_{k,i} \\ L_{i+m,i} = a_{i+m,i} - \sum_{k=1}^{i-1} L_{i+m,k} \sigma_k U_{k,i} \end{array} \right.$$

$$(5.11) \left\{ \begin{array}{l} U_{i,i+1} = a_{i,i+1} - \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i+1} \\ U_{i,i+2} = a_{i,i+2} - \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i+2} \\ U_{i,i+m-2} = a_{i,i+m-2} - \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i+m-2} \\ U_{i,i+m-1} = a_{i,i+m-1} - \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i+m-1} \\ U_{i,i+m} = a_{i,i+m} - \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i+m} \\ L_{i,i} = U_{i,i} = \frac{1}{\sigma_i} = (\alpha+1) a_{i,i} - \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i} \end{array} \right.$$

Or l'ensemble P_3 impose :

$$\left\{ \begin{array}{l} \sum_{k=1}^{i-1} L_{i+1,k} \sigma_k U_{k,i} = L_{i+1,i-1} \sigma_{i-1} U_{i-1,i} \\ \quad + L_{i+1,i-m+2} \sigma_{i-m+2} U_{i-m+2,i} \\ \quad + L_{i+1,i-m+1} \sigma_{i-m+1} U_{i-m+1,i} \\ \sum_{k=1}^{i-1} L_{i+2,k} \sigma_k U_{k,i} = L_{i+2,i-m+2} \sigma_{i-m+2} U_{i-m+2,i} \\ \sum_{k=1}^{i-1} L_{i+m-2,k} \sigma_k U_{k,i} = L_{i+m-2,i-1} \sigma_{i-1} U_{i-1,i} \\ \quad + L_{i+m-2,i-2} \sigma_{i-2} U_{i-2,i} \\ \sum_{k=1}^{i-1} L_{i+m-1,k} \sigma_k U_{k,i} = L_{i+m-1,i-1} \sigma_{i-1} U_{i-1,i} \\ \sum_{k=1}^{i-1} L_{i+m,k} \sigma_k U_{k,i} = 0 \\ \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i+1} = L_{i,i-1} \sigma_{i-1} U_{i-1,i+1} \\ \quad + L_{i,i-m+2} \sigma_{i-m+2} U_{i-m+2,i+1} \\ \quad + L_{i,i-m+1} \sigma_{i-m+1} U_{i-m+1,i+1} \end{array} \right.$$

$$(5.12) \left\{ \begin{array}{l} \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i+2} = L_{i,i-m+2} \sigma_{i-m+2} U_{i-m+2,i+2} \\ \\ \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i+m-2} = L_{i,i-1} \sigma_{i-1} U_{i-1,i+m-2} \\ \quad + L_{i,i-2} \sigma_{i-2} U_{i-2,i+m-2} \\ \\ \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i+m-1} = L_{i,i-1} \sigma_{i-1} U_{i-1,i+m-1} \\ \\ \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i+m} = 0 \\ \\ \sum_{k=1}^{i-1} L_{i,k} \sigma_k U_{k,i} = L_{i,i-1} \sigma_{i-1} U_{i-1,i} \\ \quad + L_{i,i-2} \sigma_{i-2} U_{i-2,i} \\ \quad + L_{i,i-m+2} \sigma_{i-m+2} U_{i-m+2,i} \\ \quad + L_{i,i-m+1} \sigma_{i-m+1} U_{i-m+1,i} \\ \quad + L_{i,i-m} \sigma_{i-m} U_{i-m,i} \end{array} \right.$$

On posera :

$$(5.13) \left\{ \begin{array}{l} L_{i+1,i} = \tilde{b}_i \\ L_{i+2,i} = e_i \\ L_{i+m-2,i} = f_i \\ L_{i+m-1,i} = h_i \end{array} \right\} \left\{ \begin{array}{l} U_{i,i+1} = \tilde{b}'_i \\ U_{i,i+2} = \tilde{e}_i \\ U_{i,i+m-2} = \tilde{f}_i \\ U_{i,i+m-1} = \tilde{h}_i \end{array} \right\} \left\{ \begin{array}{l} a_{i+1,i} = b_i \\ a_{i,i+1} = b'_i \\ a_{i+m,i} = c_i \\ a_{i,i+m} = c'_i \text{ et } a_{i,i} = a_i \end{array} \right.$$

Or on sait que :

$$\left\{ \begin{array}{l} a_{i+2,i} = 0 \\ a_{i+m-2,i} = 0 \\ a_{i+m-1,i} = 0 \end{array} \right\} \left\{ \begin{array}{l} a_{i,i+2} = 0 \\ a_{i,i+m-2} = 0 \\ a_{i,i+m-1} = 0 \end{array} \right.$$

En adoptant les notations (5.13) et en appliquant les formules (5.12), les formules (5.11) de cette décomposition deviennent :

$$\begin{aligned} L_{i,i} = U_{i,i} = \frac{1}{\sigma_i} = & (\alpha+1) a_i - \tilde{b}_{i-1} \sigma_{i-1} \tilde{b}'_{i-1} - e_{i-2} \sigma_{i-2} \tilde{e}_{i-2} \\ & - f_{i-m+2} \sigma_{i-m+2} \tilde{f}_{i-m+2} \\ & - h_{i-m+1} \sigma_{i-m+1} \tilde{h}_{i-m+1} - c_{i-m} \sigma_{i-m} c'_{i-m}. \end{aligned}$$

Pour la matrice L :

$$\begin{aligned} L_{i+1,i} = \tilde{b}_i = & b_i - e_{i-1} \sigma_{i-1} \tilde{b}'_{i-1} - h_{i-m+2} \sigma_{i-m+2} \tilde{f}_{i-m+2} \\ & - c_{i-m+1} \sigma_{i-m+1} \tilde{h}_{i-m+1} \\ L_{i+2,i} = e_i = & -c_{i-m+2} \sigma_{i-m+2} \tilde{f}_{i-m+2} \\ L_{i+m-2,i} = f_i = & -h_{i-1} \sigma_{i-1} \tilde{b}'_{i-1} - c_{i-2} \sigma_{i-2} \tilde{e}_{i-2} \\ L_{i+m-1,i} = h_i = & -c_{i-1} \sigma_{i-1} \tilde{b}'_{i-1} \\ L_{i+m,i} = & c_i \end{aligned}$$

Pour la matrice U :

$$\begin{aligned} U_{i,i+1} = \tilde{b}'_i = & b'_i - \tilde{b}_{i-1} \sigma_{i-1} \tilde{e}_{i-1} - f_{i-m+2} \sigma_{i-m+2} \tilde{h}_{i-m+2} \\ & - h_{i-m+1} \sigma_{i-m+1} c'_{i-m+1} \\ U_{i,i+2} = \tilde{e}_i = & -f_{i-m+2} \sigma_{i-m+2} c'_{i-m+2} \\ U_{i,i+m-2} = \tilde{f}_i = & -\tilde{b}_{i-1} \sigma_{i-1} \tilde{h}_{i-1} - e_{i-2} \sigma_{i-2} c'_{i-2} \\ U_{i,i+m-1} = \tilde{h}_i = & -\tilde{b}_{i-1} \sigma_{i-1} c'_{i-1} \\ U_{i,i+m} = & c'_i. \end{aligned}$$

Remarques :

- La décomposition incomplète DI(3) demandera plus de calcul en programmation et plus de places en mémoires que la décomposition incomplète DI(0).
- La matrice $L \Sigma U$ donnée par la décomposition DI(3) approchera la matrice A beaucoup plus que la matrice donnée par la décomposition DI(0).

11.3 - Méthodes du gradient conjugué :

* Pour les décompositions incomplètes symétriques IC(.) on utilisera le choix des matrices H et K suivant :

$$\begin{cases} H = (A A^T)^{-1} \\ K = A^T (L \Sigma L^T)^{-1} (L \Sigma L^T)^{-1} A. \end{cases}$$

Par ce choix, les deux matrices H et K sont symétriques et définies positives et la méthode (5.7) de la méthode du gradient conjugué devient :

$$\left\{ \begin{array}{l} x_0 \text{ vecteur arbitraire,} \quad r_0 = b - A x_0 \\ p_0 = A^T (L \Sigma L^T)^{-1} (L \Sigma L^T)^{-1} r_0 / \|(L \Sigma L^T)^{-1} r_0\|^2 \\ \text{et pour } i \geq 0 \\ c_i = \frac{1}{\|p_i\|^2} \\ x_{i+1} = x_i + c_i p_i \\ r_{i+1} = r_i - c_i A p_i \\ a_{i+1} = \frac{1}{\|(L \Sigma L^T)^{-1} r_{i+1}\|^2} \\ p_{i+1} = a_{i+1} A^T (L \Sigma L^T)^{-1} (L \Sigma L^T)^{-1} r_{i+1} + p_i \end{array} \right.$$

et si on pose $(L \Sigma L^T)^{-1} r_i = s_i$ cet algorithme devient :

$$\left\{ \begin{array}{l} x_0 \text{ vecteur quelconque} \quad r_0 = b - A x_0 \quad s_0 = (L \Sigma L^T)^{-1} r_0 \\ p_0 = A^T (L \Sigma L^T)^{-1} s_0 / \|s_0\|^2 \quad \text{et pour } i \geq 0 \\ c_i = \frac{1}{\|p_i\|^2} \\ x_{i+1} = x_i + c_i p_i \\ s_{i+1} = s_i - c_i (L \Sigma L^T)^{-1} A p_i \end{array} \right.$$

$$\begin{cases} a_{i+1} = \frac{1}{\|s_{i+1}\|^2} \\ p_{i+1} = a_{i+1} A^T (L \Sigma L^T)^{-1} s_{i+1} + p_i. \end{cases}$$

Ces méthodes seront notées ICCG (.).

* Pour les décompositions incomplètes non symétriques DI(.) on fera le choix suivant :

$$\begin{cases} H = (AA^T)^{-1} \\ K = A^T (L \Sigma U)^{-T} (L \Sigma U)^{-1} A. \end{cases}$$

Par ce choix, l'algorithme donnant la méthode du gradient conjugué est le suivant :

$$(5.14) \quad \begin{cases} x_0 \text{ quelconque} & r_0 = b - Ax_0 \\ p_0 = A^T (L \Sigma U)^{-T} (L \Sigma U)^{-1} r_0 / \|(L \Sigma U)^{-1} r_0\|^2 \\ \text{et pour } i \geq 0 \\ c_i = \frac{1}{\|p_i\|^2} \\ x_{i+1} = x_i + c_i p_i \\ r_{i+1} = r_i - c_i A p_i \\ a_{i+1} = \frac{1}{\|(L \Sigma U)^{-1} r_{i+1}\|^2} \\ p_{i+1} = a_{i+1} A^T (L \Sigma U)^{-T} (L \Sigma U)^{-1} r_{i+1} + p_i. \end{cases}$$

Ces méthodes seront notées DICG (.).

Remarques :

- On peut utiliser, pour l'algorithme (5.14), la transformation suivante :

$$(L \Sigma U)^{-1} r_{i+1} = s_{i+1}.$$

- Pour les deux choix on minimise :

$$E(x_i) = \|x_i - x\|^2 = (x_i - x, x_i - x).$$

- Le choix de H et K n'a pas varié pour les deux algorithmes mais les décompositions ont changé, donc on peut comparer les résultats de ces deux algorithmes comme résultats du choix des décompositions incomplètes.

- On a une majoration de l'erreur :

$$E(x_i) = \|x_i - x\|^2 \leq \|x_0 - x\|^2 \left[\frac{\lambda_{\max}(T) - \lambda_{\min}(T)}{\lambda_{\max}(T) + \lambda_{\min}(T)} \right]^{2i}$$

où :

$$* T = A^T (L \Sigma L^T)^{-1} (L \Sigma L^T)^{-1} A$$

pour les décompositions incomplètes IC(.).

$$* T = A^T (L \Sigma U)^{-T} (L \Sigma U)^{-1} A$$

pour les décompositions incomplètes DI(.).

Tableau II, 1 :

Ce tableau représente les résultats de la méthode du gradient conjugué combinée avec les décompositions incomplètes symétriques IC(.) pour le choix suivant :

$$\begin{cases} H = (AA^T)^{-1} \\ K = A^T(L \Sigma L^T)^{-1}(L \Sigma L^T)^{-1}A. \end{cases}$$

Elle sera notée ICCG (.)

$m = 19,$ $\omega = 400,$ $f = 10$ et $\epsilon = 10^{-6}$

	ICCG (0)		ICCG (1)		ICCG (3)	
Alpha	IT	$\text{Log}_{10} \ b - Ax\ $	IT	$\text{Log}_{10} \ b - Ax\ $	IT	$\text{Log}_{10} \ b - Ax\ $
0.025	14	-6.180	12	-6.166	11	-6.311
0.020	14	-6.195	12	-6.172	11	-6.334
0.015	14	-6.211	12	-6.207	11	-6.345
0.010	14	-6.226	12	-6.297	10	-6.037
0.005	14	-6.242	12	-6.411	10	-6.062
0.000	14	-6.257	12	-6.481	10	-6.067
-0.005	14	-6.274	11	-6.000	10	-6.090
-0.010	14	-6.293	11	-6.029	10	-6.158
-0.015	14	-6.318	11	-6.077	10	-6.219
-0.020	14	-6.350	11	-6.137	10	-6.214
-0.025	14	-6.390	11	-6.195	10	-6.175

Tableau II.2 :

Méthode du gradient conjugué avec les décompositions incomplètes à positions symétriques DI(.) pour le choix suivant :

$$\begin{cases} H = (A A^T)^{-1} \\ K = A^T (L \Sigma U)^{-T} (L \Sigma U)^{-1} A. \end{cases}$$

Elle sera notée : DICG (.)

$m = 19,$ $\omega = 400,$ $f = 10$ et $\epsilon = 10^{-6}$

	DICG (0)		DICG (3)	
Alpha	IT	$\text{Log}_{10} \ r\ $	IT	$\text{Log}_{10} \ r\ $
0.025	10	-6.027	5	-6.254
0.020	10	-6.096	5	-6.575
0.015	10	-6.165	5	-6.975
0.010	10	-6.237	5	-7.493
0.005	10	-6.308	4	-6.483
0.000	10	-6.381	4	-6.912
-0.005	10	-6.453	4	-7.111
-0.010	10	-6.526	4	-6.557
-0.015	10	-6.602	4	-6.144
-0.020	9	-6.034	5	-6.042
-0.025	9	-6.111	5	-6.601

II.4 - Méthodes d'orthogonalisation - minimisation :

* Pour les décompositions incomplètes symétriques $IC(.)$ on utilise le choix des matrices H et K défini dans l'exemple précédent 1.3. Ces méthodes sont notées $ORTHOMIN\ k(.)$, k nombre d'orthogonalisation prend les valeurs 1 et 2 ; pour des valeurs de k supérieures à 2 l'algorithme converge pour une précision égale à 10^{-5} . Les résultats de ces méthodes sont présentés dans le tableau II.2.

* Pour les décompositions incomplètes à "positions symétriques", on fait le choix suivant :

$$\begin{cases} H = I \\ K = (L \Sigma U)^{-1} A^{-T}. \end{cases}$$

L'algorithme devient :

$$(5.15) \quad \left\{ \begin{array}{l} k \text{ nombre d'orthogonalisations} \\ x_0 \text{ quelconque} \quad r_0 = b - Ax_0 \quad p_0 = (L \Sigma U)^{-1} r_0 \\ \text{pour } i \geq 0 \\ c_i = (r_i, A(L \Sigma U)^{-1} r_i) / \|A p_i\|^2 \\ x_{i+1} = x_i + c_i p_i \\ r_{i+1} = r_i - c_i A p_i \\ a_{i-j}^{i+1} = (A(L \Sigma U)^{-1} r_{i+1}, A p_{i-j}) / \|A p_{i-j}\| \quad j=0, \dots, k-1 \\ p_{i+1} = (L \Sigma U)^{-1} r_{i+1} - \sum_{j=0}^{k-1} a_{i-j}^{i+1} p_{i-j}. \end{array} \right.$$

Cet algorithme converge si $K = (L \Sigma U)^{-1} A^{-T}$ est définie positive. Or $(L \Sigma U)^{-1}$ approche la matrice A^{-1} .

$$\begin{aligned} K &= (L \Sigma U)^{-1} A^{-T} \\ &= (L \Sigma U)^{-1} A A^{-1} A^{-T} \\ &\approx A^{-1} A^{-T} \end{aligned}$$

donc la matrice reste proche d'une matrice symétrique et définie positive,

Ces méthodes sont notées ORTHODI (.).

Les résultats sont présentés dans le tableau II.3.

Tableau II.3 :

Ce tableau représente les résultats de la méthode d'orthogonalisation - minimisation combinée avec les décompositions incomplètes symétriques IC(.) pour le choix suivant :

$$\begin{cases} H = I \\ K = (L \Sigma L^T)^{-1} A^{-T}. \end{cases}$$

Elle sera notée ORTHOMIN k(.) : k représentera le nombre d'orthogonalisations

$m = 19,$ $\omega = 400,$ $f = 10$ et $\epsilon = 10^{-6}$

Alpha	ORTHOMIN1(0)		ORTHOMIN1(1)		ORTHOMIN1(3)		ORTHOMIN2(0)		ORTHOMIN2(1)		ORTHOMIN2(3)	
	IT	$\log_{10} r $	IT	$\log_{10} r $	IT	$\log_{10} r $	IT	$\log_{10} r $	IT	$\log_{10} r $	IT	$\log_{10} r $
0.025	9	-6.373	7	-6.325	7	-6.604	10	-6.463	8	-6.349	8	-6.617
0.020	9	-6.384	7	-6.409	7	-6.760	9	-6.510	7	-6.409	7	-6.729
0.015	9	-6.395	7	-6.492	7	-6.918	9	-6.537	7	-6.491	7	-6.935
0.010	9	-6.404	7	-6.564	7	-7.073	9	-6.555	7	-6.572	7	-7.163
0.005	9	-6.412	7	-6.662	6	-6.132	9	-6.577	7	-6.649	6	-6.193
0.000	9	-6.419	7	-6.696	6	-6.592	8	-6.002	7	-6.702	6	-6.553
-0.005	9	-6.426	7	-6.677	6	-6.306	8	-6.016	7	-6.703	6	-6.697
-0.010	9	-6.433	7	-6.800	6	-6.172	8	-6.026	6	-6.073	6	-6.675
-0.015	9	-6.439	7	-6.797	7	-7.136	8	-6.039	7	-6.819	6	-6.358
-0.020	9	-6.445	7	-6.805	7	-6.541	8	-6.054	7	-6.813	6	-6.023
-0.025	9	-6.452	7	-6.808	7	-6.105	8	-6.070	7	-6.790	7	-6.831

Tableau II.4 :

Méthode d'orthogonalisation - minimisation avec les décompositions incomplètes à positions symétriques pour le choix suivant :

$$\begin{cases} H = I \\ K = (L \Sigma U)^{-1} A^{-T}. \end{cases}$$

Elle sera notée : ORTHODI k(.)

$m = 19$, $\omega = 400$, $f = 10$ et $\epsilon = 10^{-6}$

Alpha	ORTHODI 1(0)		ORTHODI 1(3)		ORTHODI 2(0)		ORTHODI 2(3)	
	IT	$\log_{10} \ r\ $	IT	$\log_{10} \ r\ $	IT	$\log_{10} \ r\ $	IT	$\log_{10} \ r\ $
0.025	6	-6.524	4	-7.199	6	-6.520	4	-7.077
0.020	6	-6.552	4	-7.362	6	-6.555	4	-7.174
0.015	6	-6.580	3	-6.005	6	-6.542	3	-6.014
0.010	6	-6.608	3	-6.240	6	-6.616	4	-7.809
0.005	6	-6.638	3	-6.438	6	-6.536	4	-7.052
0.000	6	-6.666	3	-6.497	6	-6.630	3	-6.294
-0.005	6	-6.696	3	-6.455	6	-6.643	4	-6.472
-0.010	6	-6.723	3	-6.241	6	-6.739	4	-6.540
-0.015	6	-6.750	3	-6.044	6	-6.703	4	-7.591
-0.020	6	-6.789	4	-7.604	6	-6.766	4	-7.585
-0.025	6	-6.822	4	-7.328	6	-6.764	4	-6.960

11.5 - Méthode du double-gradient conjugué :

Cette méthode est utilisée, dans ce chapitre, avec les décompositions incomplètes à "positions symétriques" $DI(.)$ définies dans 11.2.2.

Au lieu de résoudre le système $Ax = b$ on le transforme :

$$(L \Sigma U)^{-1} Ax = (L \Sigma U)^{-1} b$$

et on applique l'algorithme (4.12) à ce dernier

$$(5.16) \quad \left\{ \begin{array}{l} x_0 \text{ un vecteur quelconque,} \\ r_0 = \bar{r}_0 = p_0 = \bar{p}_0 = (L \Sigma U)^{-1} [b - Ax_0] \quad \text{pour } i \geq 0 \\ \alpha_i = (\bar{r}_i, r_i) / (\bar{p}_i, (L \Sigma U)^{-1} A p_i) \\ x_{i+1} = x_i + \alpha_i p_i \\ r_{i+1} = r_i - \alpha_i (L \Sigma U)^{-1} A p_i \\ \bar{r}_{i+1} = \bar{r}_i - \alpha_i A^T (L \Sigma U)^{-T} \bar{p}_i \\ \beta_i = (\bar{r}_{i+1}, r_{i+1}) / (\bar{r}_i, r_i) \\ p_{i+1} = r_{i+1} + \beta_i p_i \\ \bar{p}_{i+1} = \bar{r}_{i+1} + \beta_i \bar{p}_i. \end{array} \right.$$

Cet algorithme converge si la matrice symétrique :

$$\tilde{A} = A^T (L \Sigma U)^{-T} + A (L \Sigma U)^{-1} \text{ est définie positive.}$$

On remarque que la matrice $L \Sigma U$ approche la matrice A . Donc :

$$\tilde{A} \approx A^T \tilde{A}^{-T} + A \tilde{A}^{-1}$$

et par conséquent, la matrice $\tilde{A}$ approche une matrice qui est définie positive.

Tableau II, 5 :

Méthode du double-gradient conjugué avec les décompositions incomplètes à positions symétriques DI (.).

Elle sera notée DIDCG (.)

$m = 19,$ $\omega = 400,$ $f = 10$ et $\epsilon = 10^{-6}$

	DIDCG (0)		DIDCG (3)	
Alpha	IT	$\text{Log}_{10} \ r\ $	IT	$\text{Log}_{10} \ r\ $
0.025	5	-6.051	4	-7.688
0.020	5	-6.079	4	-7.819
0.015	5	-6.103	4	-7.917
0.010	5	-6.128	3	-6.334
0.005	5	-6.154	3	-6.732
0.000	5	-6.618	3	-6.754
- 0.005	5	-6.207	3	-6.742
- 0.010	5	-6.235	3	-6.510
- 0.015	5	-6.263	3	-6.316
- 0.020	5	-6.291	3	-6.121
- 0.025	5	-6.318	4	-7.516

Conclusions : pour un problème non symétrique

- * La méthode d'orthogonalisation - minimisation et la méthode du double-gradient conjugué préconditionné convergent mieux que la méthode du gradient conjugué préconditionné.
- * La méthode du double-gradient conjugué préconditionné demande plus de calcul et plus de place en mémoire que la méthode d'orthogonalisation - minimisation et les deux méthodes convergent pratiquement en même nombre d'itérations.

III - TROISIEME EXEMPLE

La matrice A n'est pas une H-matrice mais elle est anti-symétrique et définie positive.

$$A = \begin{bmatrix} B & 2I & \\ -2I & B & 2I \\ & -2I & B \end{bmatrix}$$

avec :

$$B = \begin{bmatrix} 2 & & & & \\ & 1 & & & \\ & & -1 & & \\ & & & -1 & \\ & & & & 2 \end{bmatrix}$$

Pour cet exemple, on utilise les décompositions incomplètes à "positions symétriques" DI(.) définies dans II.2 et données par les formules (5.9) et (5.11).

On applique l'algorithme d'orthogonalisation - minimisation décrit dans l'exemple précédent pour le même choix des matrices H et K (5.15) et l'algorithme du double-gradient conjugué (5.16).

Les méthodes d'orthogonalisation - minimisation combinées avec les décompositions sont notées ORTHMIN k(.) et les résultats pour ces méthodes sont donnés par le tableau III.1.

Les méthodes du double-gradient conjugué avec les décompositions incomplètes sont notées DIDCG(.) et les résultats sont groupés dans le tableau III.2.

Tableau III.1 :

Méthode d'orthogonalisation - minimisation combinée avec les décompositions incomplètes à positions symétriques DI(.) pour le choix suivant :

$$\begin{cases} H = I \\ K = (L \Sigma U)^{-1} A^{-T} \end{cases}$$

Elle sera notée ORTHMIN k(.)

m = 19, f = 10, ε = 10⁻⁶

	ORTHMIN 1(0)		ORTHMIN 1(3)		ORTHMIN 2(0)		ORTHMIN 2(3)	
Alpha	IT	Log ₁₀ r	IT	Log ₁₀ r	IT	Log ₁₀ r	IT	Log ₁₀ r
0.025	14	-6.081	6	-7.227	13	-6.063	6	-7.175
0.020	14	-6.018	6	-7.279	13	-6.022	6	-7.253
0.015	15	-6.475	6	-7.314	14	-6.568	6	-7.228
0.010	15	-6.412	6	-7.325	14	-6.519	6	-6.819
0.005	15	-6.351	6	-7.308	14	-6.468	6	-7.296
0.000	15	-6.291	6	-7.263	14	-6.418	6	-7.129
-0.005	15	-6.232	6	-7.194	14	-6.370	6	-7.244
-0.010	15	-6.173	6	-7.112	14	-6.326	6	-7.135
-0.015	15	-6.115	6	-7.030	14	-6.283	6	-7.063
-0.020	15	-6.058	6	-6.949	14	-6.244	6	-7.069
-0.025	15	-6.001	6	-6.870	14	-6.206	6	-6.946

Tableau III.2 :

Méthode du double-gradient conjugué combinée avec les décompositions incomplètes à positions symétriques DI (.).

Elle sera notée DIDCG (.)

$m = 19$, $f = 10$ et $\epsilon = 10^{-6}$

	DIDCG (0)		DIDCG (3)	
Alpha	IT	$\text{Log}_{10} \ r\ $	IT	$\text{Log}_{10} \ r\ $
0.025	14	-6.688	6	-6.710
0.020	14	-6.397	6	-7.153
0.015	14	-6.074	6	-7.317
0.010	15	-6.600	6	-7.360
0.005	15	-6.413	6	-7.301
0.000	15	-6.777	6	-7.051
-0.005	15	-6.879	7	-8.337
-0.010	15	-6.989	6	-7.139
-0.015	15	-7.060	6	-6.107
-0.020	15	-6.676	6	-6.619
-0.025	15	-6.220	6	-6.864

BIBLIOGRAPHIE

- [1] AXELSSON O. :
Solution of linear systems of equations : iterative methods
Sparse matrix techniques Copenhagen 1976
Lecture notes in maths 572 Springer-Verlag (1976).
- [2] AXELSSON O. :
Conjugate gradient type methods for unsymmetric and inconsistent systems of linear equations
Linear algebra and its applications vol. 29 (1980) pp. 1-16.
- [3] BERMAN A. et PLEMMONS R.J. :
Non negative matrices in the mathematical sciences
Academic Press (1979).
- [4] COHEN A.I. :
Rate of convergence of several conjugate gradient algorithms
SIAM, J. Numer. Anal., vol. 9, n°2, 1972, pp. 248-259.
- [5] CONCUS P. et GOLUB G.H. :
A generalized conjugate gradient method for non symmetric systems of linear equations
Report STAN - CS - 76535 Stanford University (1976).
- [6] CONCUS P., GOLUB G.H. et O'LEARY D.P. :
A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations
In Sparse matrix computations
Ed. BUNCH J.R. et ROSE D.J., Academic Press (1976),
pp. 309-332.
- [7] DANIEL J.W. :
The conjugate gradient method for linear and nonlinear operator equations
SIAM, J. Numer. Anal., vol. 4 (1967) pp. 10-26.

- [8] DOUGLAS J. et DUPONT T. :
Preconditionned conjugate gradient iteration applied to Galerkin
methods for a mildly nonlinear Dirichlet problem
in Sparse Matrix Computations
Ed. BUNCH J. R. et ROSE D. J., Academic Press (1976),
pp. 333-348.
- [9] GASTINEL N. :
Analyse numérique linéaire
Ed. Hermann, Paris 1966.
- [10] GLOWINSKY R., MANTEL B., PERIAUX J., PIRONNEAU O. et
POIRIER G. :
An efficient preconditioned conjugate gradient method applied to
linear problems in fluid dynamics via least square formulations
In Computing methods in applied sciences and engineering
Ed. Glowinsky R. et Lions J. L.
North-Holland Publishing Compagny
INRIA 1980, pp. 444-472.
- [11] HESTENS M. et STIEFEL E. :
Methods of conjugate gradient for solving linear systems
J. Res. Nat. Bur. of Stand. B49 (1952), pp. 409-436.
- [12] KERSHAW D. S. :
The incomplete Cholesky-conjugate gradient method for the
iterative solution of systems of linear equations
J. of Comp. Physics, vol. 26, (1978), pp. 43-65.
- [13] MANTEUFFEL T. A. :
An incomplete factorization technique for positive definite linear
systems
Math. of Comp. (1980), vol. 34, n°150, pp. 473-497.
- [14] MANTEUFFEL T. A. :
Solving structures problems iteratively with a shifted incomplete
Cholesky preconditioning
In Computing Methods in Applied Sciences and Engineering

Ed. Glowinsky R. et Lions J.L.
North-Holland Publishing Compagny
INRIA 1980, Paris.

- [15] MEIJERINK J.A. et VAN DER VORST H.A. :
An iterative solution method for linear systems of which the
coefficient matrix is a symmetric M-matrix
Math. of Comp. (1977), vol. 31, n°137, pp. 148-162.
- [16] MEIJERINK J.A. et VAN DER VORST H.A. :
Guidelines for the usage of incomplete decomposition in solving
sets of linear equations as they occur in practical problems
J. of Comp. Physics (1981), vol. 44, pp. 134-155.
- [17] MEURANT G. et GOLUB G.H. :
Résolution numérique des grands systèmes linéaires
Note CEA-N-2268
Commissariat à l'Energie Atomique
Centre d'Etudes de Limeil.
- [18] MIELLOU J. C. :
Une méthode du gradient conjugué préconditionné dans le cas
d'opérateurs non linéaires
XV° Colloque d'analyse numérique, 24-28 Mai 1982, Belgodère,
Corse.
- [19] MIELLOU J. C. :
Variantes synchrones et asynchrones de la méthode alternée
de Schwarz
Rapport de recherche L.A.N.I., 4 Mars 1982, Besançon.
- [20] PETRAVIC et KUO-PETRAVIC :
An ILUCG algorithm which minimizes in the euclidian norm
J. of Comp. Physics (1979), vol. 32, pp. 263-269.
- [21] REID J. K. :
On the method of conjugate gradient for the solution large
sparse systems of linear equations

In Large sparse sets of linear equations
Ed. Reid J. K., Academic Press (1971).

- [22] RAVIART P. A. :
Cours d'analyse numérique, chapitre 4
Ecole polytechnique, Paris (1978).
- [23] ROBERT F. :
Matrices non négatives et normes vectorielles
Polycopié Grenoble, IMAG (1972).
- [24] ROBERT Y. :
Factorisation incomplète d'une matrice définie positive (extension
au cas non symétrique)
Peprint, IMAG, Grenoble (1981).
- [25] TAKAHASI H. et NODERA T. :
New variants of the conjugate gradient algorithms
Department of Mathematics, Faculty of engineering,
KEIO UNIVERSITY - JAPAN.
- [26] VAN DER VORST H. A. :
Iterative solution methods for certain sparse linear systems
with a non symmetric matrix arising from PDE-problems
J. of Comp. Physics (1981), vol. 44, pp. 1-19.
- [27] VARGA R. S. :
M-matrix theory and recent results in numerical linear algebra
In Sparse Matrix Computations
Ed. BUNCH J. R. et ROSE D. J., Academic Press (1976),
pp. 375-385.
- [28] VARGA R. S. :
Matrix iterative analysis
Prentice Hall, Englewood Cliffs, New-Jersey 1962.
- [29] VARGA R. S., SAFF E. B. et MEHRMANN V. :
Incomplete factorization of matrices and connection with
H-matrices
SIAM, J. Numer. Anal. (1980), vol. 17, n°6, pp. 787-793.

- [30] VINSOME P.K.W. :
ORTHOMIN, an iterative method for solving sparse sets of
simultaneous linear equations
Society of Petroleum Engineers of AIME, Paper number SPE
5729 (1976)
American Institute of Mining, Metallurgical and Petroleum
Engineers, Inc.
- [31] WILKINSON J.H. :
The algebraic eigenvalue problem
Clarendon Press 1965.
- [32] JOLY P. :
Résolutions de systèmes linéaires non symétriques par des
méthodes de gradient conjugué
Publications du Laboratoire d'Analyse Numérique
Université Pierre et Marie Curie CNRS.

TABLE DES MATIERES
=====

	page
<u>INTRODUCTION</u>	1
 <u>CHAPITRE 1 : DECOMPOSITIONS INCOMPLETES DES</u> <u>M-MATRICES</u>	3
I - Définition et propriétés des M-matrices.....	4
II - Théorèmes des décompositions incomplètes des M-matrices.....	9
III - Théorèmes de convergence et de stabilité numérique.....	13
 <u>CHAPITRE 2 : DECOMPOSITIONS INCOMPLETES DES</u> <u>H-MATRICES</u>	18
I - Définition et propriétés des H-matrices.....	19
II - Théorèmes des décompositions incomplètes des H-matrices.....	28
III - Théorèmes de convergence et de stabilité numérique.....	36
IV - Théorèmes de caractérisation des H-matrices	41
 <u>CHAPITRE 3 : UNE METHODE DE DECOMPOSITIONS INCOMPLETES</u> <u>D'UNE MATRICE QUELCONQUE (NON NECESSAIRE-</u> <u>MENT UNE H-MATRICE)</u>	45
I - L'idée de la méthode.....	46
II - Cas des matrices définies positives.....	49
III - Etude de la convergence pour le cas d'une matrice définie positive.....	52

	page
<u>CHAPITRE 4 : SUR QUELQUES METHODES ITERATIVES DE RESOLUTION D'UN SYSTEME LINEAIRE.....</u>	54
I - Notations et définitions.....	55
II - La méthode d'orthogonalisation - minimisation : ORTHOMIN.....	59
III - La méthode généralisée du gradient conjugué : CG.....	72
IV - La méthode du double-gradient conjugué.....	76
 <u>CHAPITRE 5 : ESSAIS NUMERIQUES.....</u>	79
I - Premier exemple : M-matrice symétrique	81
II - Deuxième exemple : M-matrice non symétrique.....	104
III - Troisième exemple : matrice anti-symétrique et définie positive.....	124
 BIBLIOGRAPHIE	127

RÉSUMÉ

On généralise la LU-décomposition incomplète à des H-matrices et on énonce un résultat assurant la convergence de la méthode de point fixe, associée à une telle décomposition.

On donne une forme généralisée de l'algorithme d'orthogonalisation-minimisation, utilisé pour résoudre un système linéaire dans un cas particulier, et on énonce un

résultat sur les conditions de convergence de cet algorithme.

On compare, ensuite, cet algorithme à d'autres plus connus (l'algorithme du gradient conjugué et l'algorithme du double-gradient conjugué) pour des problèmes non symétriques.

MOTS CLEFS

- H-matrices
- Décompositions incomplètes
- Itérations
- Gradient conjugué
- Orthogonalisation-minimisation